

Can Japan survive a new government?

Hopes outside Japan that last month's general election will weaken Japan's technological competitiveness are misplaced; there is much that the new government can do to sustain the momentum.

When General Douglas MacArthur allowed Japan after the Second World War to adopt a parliamentary rather than a presidential system of government, he was no doubt aware of the constitutional nicety that Japan had no need of an elected head of state, having an Emperor already. But for nearly half a century, Japan has managed to wring political stability from a recipe that, in other people's hands, is a means of changing governments at regular and even irregular intervals: during 38 years of steady social and economic improvement, the Liberal Democratic Party (LDP) has held a firm grip on the government of Japan. But now it is on the way out. Last month's election produced only an inconclusive result; it is not yet clear what coalition (which may include the LDP) will form the next government. So does that mean that Japan is at a turning point in its remarkable history of success in the post-war world? Will Japan no longer be the formidable competitor to which its rivals have grown accustomed?

It is too soon for envious competitors elsewhere to be throwing their hats in the air. The fact of the election may have been a surprise, but its outcome is less so; sooner or later, the preoccupation of politicians of the ruling party with gifts of money was certain to catch up with them. But at least at the beginning, a new government, however radical its ambitions, is unlikely to alter significantly the way in which Japan does government business. And to the small extent that Japan's economic success stems from government policy, the chances are that there will be no immediate change. Toyota factories in the United States or elsewhere will not begin assembling substandard motor-cars because there is a new government in Tokyo.

Misunderstanding

To suppose otherwise is to misunderstand Japan and its powerful bureaucracy. First, Japan has very little of what is elsewhere called politics. Not for Japan great crises over questions such as President Bill Clinton's plan for universal health care (still waiting in the wings) or Britain's adoption of the Maastricht Treaty (which the British government secured last week by the skin of its teeth, and at huge political cost). A political crisis in Japan may instead consist of a row about the role of the armed services (called self-defence forces) outside Japan and the propriety of a 3 per cent sales tax. Government ministers are rarely powerful figures with the responsibility to mould a whole slew of policies, but are more often figureheads who are shown

great respect while remaining in office, but who are replaced ever year or two.

The bureaucracy has been a much more important determinant of Japan's success. Those recruited as officials are among the brightest and the best. They benefit from the continuity with which Japan provides them, and they are identifiable among their constituents outside as influential people. In technology, for example, the Ministry of International Trade and Industry (MITI) has been a powerful encouragement of innovation, but (contrary to legend) without spending much in the way of public funds. (The imaginative daring of companies has been even more important.) MITI's successes, in fields such as steel-making, motor-cars and electronics, have turned more on winning agreement to substantial programmes of research among the groups of the companies concerned. It is unthinkable that a new government will abandon this way of working.

Finance

But not everything that MITI has touched has turned to gold, which points in an interesting direction. The widely advertised 'fifth generation computer project', for example, floundered from the beginning. It was misconceived. It would probably not have been begun at all if MITI had not been required by the Ministry of Finance to label its encouragement of electronic technology in a socially identifiable way. MITI, in short, may be powerful, but the Ministry of Finance is much more so. And the same is true of the role of the Ministry of Finance in relation to other responsibilities for science and technology. Its influence explains the slavish egalitarianism of the education ministry in relation to the national universities as well as the continuing difficulties of supporting basic research in them, for example.

So it is no wonder that some of the new parties in Japan's political spectrum are gunning for the Ministry of Finance. Tsomoto Hata, the leader of the Shinseiko Party splintered from the LDP and a possible successor as prime minister to Kiichi Miyazawa, asks openly that the ministry should return its budget-making powers to the cabinet and to elected politicians. Morihiro Hosokawa, leader of the Japan New Party, echoes that. Constitutionally, they are correct. What is democracy for if, in the end, unelected officials decide what can or even should be done? But a government likely to have to seek a new election in the near



future would be ill-advised to push for too much change too soon, even if the finance ministry is as culpable as supposed. So what else is to be done?

Luckily, where science and technology are concerned, the machinery for getting a grip on government exists; it needs to be made effective. For many years, Japan has maintained a Council on Science and Technology (CST) which, by including the prime minister, the minister of finance, other relevant ministers and several people from industry and academic life, is a cross between an advisory council and an executive committee of the cabinet (see *Nature* **359**, 579; 1992). One of its weaknesses is that it is regarded as a creature of the education ministry (which means that other ministries regard it with profound suspicion). But a new government could change that, as it could also change the council's over-cumbersome way of working. In short, without fighting the Ministry of Finance head-on, a new government in Tokyo could, with a little courage, bring about an overdue reform and ensure that Japan's flair for technical innovation is not dissipated.

That would be beneficial for Japan in many ways. Over several years, a consensus has emerged on the need to strengthen basic research at the universities. Not only academics but industrial associations are saying just that; Japan is beginning to sense a scarcity of skill. Last year, the Council on Science and Technology declared that the budgets in the field should be doubled, but characteristically neglected to say by when that state of grace should be attained.

That is a cause the new government, whatever form it takes, could usefully take up. The issue is literally that of whether Japan can sustain the momentum of the past four decades. The new government it will know that Japan's voters will be watching like hawks to see that its arrival does not make Japan falter. There is little comfort in that for Japan's rivals elsewhere. □

The lab of last resort

The British government should give a better explanation of its decision to manage a public laboratory directly.

QUITE what the British government hopes will emerge from its decision to take over the direct management of what is now the largest public laboratory in Britain will remain a mystery for some time (see page 369). The Rutherford Appleton Laboratory, itself the residual legatee of many past research programmes, now seems destined to become the only tangible residuum of an age of relative innocence in the administration of research in Britain. The history of the laboratory tells the story.

At the outset, the Rutherford Laboratory was founded (just across the wire fence from the Atomic Energy Research Establishment at Harwell in Oxfordshire) as Britain's premier high-energy physics laboratory. But the proton accelerator called Nimrod was too little and too late. The research councils' commitment to the venture was

nevertheless confirmed by the physical transfer to the site of the Radio Research Station, otherwise the Appleton Laboratory, from its previous location at Slough, some 30 miles to the east.

With the passage of time, the laboratory became a provider of research facilities generally to the academic research community. If there was a satellite project to be managed, the laboratory would take it on, as it took on the management of the complicated software project to transfer images of the sky from one academic laboratory to another. Latterly, it has become the chief British source of neutrons for research; when the Daresbury laboratory is merged, the Rutherford Appleton Laboratory (RAL) will be the British source of synchrotron radiation as well.

There is no doubt of the competence of RAL, whose excellent staff have an enviable record of versatility and achievement. The question that seems not to have been asked during the long preparation of the British government's white paper on research is whether there is a need for a laboratory of this kind whose cost is likely to be a substantial fraction, perhaps a third, of what the new research councils are able to spend on research in academic institutions.

If the question had been asked, it might have been answered in ways that would disappoint RAL. For the past quarter of a century, it has not been easy to understand why the support of substantial research projects should always be managed by this central laboratory rather than by their prime movers, the researchers winning financial support for them. True, in fields in which many researchers are active, central support may bring economies, but there are not many of them in Britain. And when one of the complaints against British basic science is that it is not sufficiently in touch with the real world, would it not be an advantage that those who win support for ambitious projects should also have to cut their teeth on their effective management and execution? On the face of things, RAL's present role works against the government's declared intention of making British science more aware of the realities of industrial life.

There are further dangers in the arrangements now planned. RAL is to be self-standing in the sense that it will earn its keep from the reimbursement of the costs of research support provided by the research councils. But what if, for example, under the new pattern of next year and beyond, the research councils commission no research in some of the fields RAL is ready to support? Will the costs of those unrequited services be spread over the costs of research in other fields? Or will RAL become the researcher of last resort when its own people use equipment unwanted by outsiders to throw back the frontiers of ignorance in fields that, of necessity, will be determined historically and not by new foresight methods on which the government is pinning its faith? It will be interesting, in these ambiguous circumstances, to see how the government manages to write its promised "mission statement" for RAL. □

Rutherford and Daresbury labs to merge as 'self-standing institution'

London. Britain's largest academic-support laboratory, the Rutherford Appleton Laboratory (RAL), with a budget of £58 million a year, will in future be financed directly by the Office of Science and Technology.

At the same time, RAL will be merged with the Daresbury Laboratory, best known for its synchrotron radiation source, to which British and many other research groups have access.

Both laboratories have hitherto been run by the Science and Engineering Research Council (SERC). RAL is near Didcot in Oxfordshire; Daresbury is 150 miles north in Cheshire. Under the new arrangement, the merged laboratory will have "self-standing" status under the Office of Public Service and Science.

The move, announced last week by science minister William Waldegrave, forms part of the reorganization of the research councils proposed by the government in a white paper (policy document) in May, and in particular the division of the SERC into an Engineering and Physical Sciences Research Council (EPSRC) and a Particle Physics and Astronomy Research Council (PPARC).

Both laboratories started as spin-offs from the government's interest after the Second World War in nuclear energy, and at first focused on nuclear research. They have since evolved into multidisciplinary centres housing large research equipment and facilities used by university scientists from Britain and abroad.

Rutherford — at which Britain's early particle accelerators were built — now operates the spallation neutron source ISIS, and several other major engineering and supercomputer facilities. Daresbury hopes to persuade the government to back plans for a new international synchrotron radiation facility, following premature closure of its Nuclear Structure Facility earlier this year (see *Nature* 362, 278; 1993).

One proposal made during discussions leading up to the white paper was that each laboratory should be attached to one of the new research councils. But Waldegrave has accepted the view of Sir David Phillips, chairman of the Advisory Board for the Research Councils, that this would be impractical because the laboratories are multidisciplinary.

Under the new arrangements, a single director based at Rutherford would be responsible for both institutions (a description of Daresbury in Phillips's draft conclusions as an "outpost" of Rutherford has been dropped in the final version of his advice to Waldegrave). The laboratories will share a

common management and a joint mission statement. This will outline what services they are expected to provide the research councils, but also allow them to bid for research and service contracts from other public and private organizations.

Staff at the laboratories have generally welcomed the government's decision to let the new organization stand on its own. "The shorter management chain will mean we will be able to make our activities leaner, cleaner and more effective," says Alan Leadbetter, director of Daresbury.

Some are concerned, however, that their new status might be a first step towards privatization — a fate that is being considered for all public laboratories following publication of the white paper, and an accompanying report prepared by the government's efficiency adviser Sir Peter Levene and its chief scientific adviser Bill Stewart.

Staff would strongly resist any such move as inappropriate because the institutions depend heavily on public funds. "We are very concerned about the possibility of privatization, since we do not feel that either institution could be run on a profit basis," says Tony Bell of the Institute of Professionals, Managers and Specialists, the

union that represents scientific staff at the two laboratories.

Senior laboratory officials say that, apart from a proposal that the new organization could eventually be turned into an independent not-for-profit company, full privatization has not been seriously discussed: it would for example create many practical difficulties over ownership of research equipment.

Leadbetter also dismisses concerns that the laboratories will be more vulnerable to future budget cuts now that they have no research council to protect them. "Look at what happened [to the Nuclear Structure Facility] under the old system," he says. "But we are worried about the prospects of reduced research budgets, tighter management, and the government's greater willingness to close public bodies than in the past."

Waldegrave also announced last week that he had accepted Phillips's proposals to transfer most SERC-funded bioengineering research to the new Biotechnology and Biological Sciences Research Council, and not to the EPSRC as the SERC had hoped (see *Nature* 364, 272; 1993).

David Dickson

Aspirin on trial as HIV treatment

Washington. An \$88,000 clinical trial at a New York hospital is to test whether aspirin is useful in treating HIV infection.

The nonprofit Community Research Initiative on AIDS will fund the trial: no public money is involved. Dr Donald Kotler of St Luke's-Roosevelt Hospital Center will lead it. Beginning in August, he will treat 46 asymptomatic HIV-infected volunteers for eight weeks with either aspirin or a placebo; the aspirin dose of four grams a day is similar to that commonly used to treat severe arthritis.

Kotler, who has worked with people suffering from AIDS in New York since the epidemic began a decade ago, says he has test-tube evidence that salicylates — a class of compounds of which aspirin is the best known — can inhibit HIV. He is ready for some ridicule over the use of the world's most famous catch-all medicine in the trial, but says: "I'm more concerned that some people might read about this and start popping aspirins." As for peer scepticism, he says: "I'm not sure who is in a position to start throwing stones."

Dr David Ho, director of the Aaron Diamond AIDS Research Center in New York,

will help to assess the outcome of the study, without necessarily endorsing Kotler's hypothesis. "I don't know much about the protocol," he says. "I know a little bit about Kotler's rationale and he has some preliminary data. This is an easy thing to do and is the right type of project for a community-based organization. Mainstream researchers are going to be sceptical, but if he has some data it should be tested out."

Kotler thinks that HIV pathogenesis involves a vicious circle of inflammation and infection in the intestine that treatment by salicylates might break. He has gathered *in vitro* and *in vivo* data to support his theory, but says he has not attracted the funding to produce as much "hard data" as he would like.

Some of the more promising data have involved salicylates other than aspirin. But he will use aspirin because if such a cheap therapy worked it would offer hope to so many people. No-one, least of all Kotler, is touting aspirin as the next wonder-cure for AIDS: the best hope is that it will buy some patients a little more time.

Colin Macilwain

Japanese go it alone with earthquake prediction

Tokyo. A new five-year plan for earthquake prediction is expected to be approved tomorrow (30 July) by a committee of Japan's Ministry of Education, Science and Culture (MESC). But despite more than 18 months of controversy, it differs little from the previous plan.

Japan's earthquake prediction programme was launched in the 1960s, and has cost more than \$1 billion and involved more than 500 scientists. Its credibility was shaken last year, when scientists — some working on the programme — questioned its underlying principles. They also criticized the fact that the programme was reviewed by its founders (see *Nature* 356, 464; 1992).

In May, six independent experts reviewed an internal report on the present six-year plan which ends next March. These included Masuo Ida, a former official of the education ministry and a vocal critic of the programme, Saburo Nagakura, president of the Graduate University for Advanced Studies, and heads of the volcanological and seismological societies of Japan.

The next five-year plan was reviewed a few weeks ago by outside experts, but they did not include the most vocal critics from the previous review. The final plan incorporates none of the criticisms made in the latest closed external review.

When Japan started its programme, many scientists worldwide believed that earthquakes could be predicted by collecting vast amounts of empirical data and sifting these for "precursors". The United States has now abandoned all but a small prediction programme in California, and spends most money and effort on disaster mitigation. But leaders of Japan's programme insist they will soon be able to predict earthquakes.

One of the plan's few new elements is a call for research on "evaluation of the potential of earthquake occurrence". This seems to mean assessment of where earthquake-prone regions are within "the cycle" of major earthquake events. This idea has been around for years. Critics say it is flawed because earthquakes do not occur in regular cycles, and it is impossible to position a region within a cycle beyond making generalized forecasts based on plate tectonics.

Members of the drafting committee who opposed this call were overruled by programme founders. One scientist close to the drafting process says the reason seems to be "to continue to trick the Japanese public into thinking earthquake prediction is possible."

What is new is a call for more ocean-bottom observations: more than four-fifths of earthquakes in Japan occur under the sea. But the report does not say where the new ocean-bottom stations will be. Scientists are

divided on the matter: one group favours a site at the junction of the Kurile and Japan trenches off Hokkaido island in the North, while another, led by Tomowoo Hirose of Tohoku University, wants them around a platform in a gas field off the coast of Fukushima Prefecture near the university.

Although two major earthquakes have hit the Japan Sea over the past decade with serious loss of life, the report does not define new areas where observations should be intensified or define new regions for intensified earthquake observation. After an earthquake hit the Japan Sea two weeks ago, Kiyoo Mogi, head of the coordinating com-

Britain to support aerospace

London. Aerospace is likely to become one of the first industrial sectors to benefit from the British government's new-found enthusiasm for aligning science funding with a consensus view of industry's long-term needs.

Tim Sainsbury, minister for industry, announced last week that the Department of Trade and Industry (DTI) will discuss with other government departments how to identify the priority needs of the aerospace industry, and how to meet them. The public consultation will include those with responsibility for publicly funded aviation research programmes in bodies ranging from the Science and Engineering Research Council (SERC) to the Ministry of Defence.

The move follows claims from the aerospace industry that a collapse in military and civilian orders has undermined its capacity to invest in long-term research. Such research, it says, is essential to its long-term prosperity (see *Nature* 362, 484; 1993). It is also consistent with the government's future approach to research funding outlined in the recent white paper (policy document).

The House of Commons Select Committee on Science and Technology last week backed industry's call that the government should support research in key technologies. The committee claimed that the United States and other European governments — in particular France and Germany — spend much more than Britain on backing their aerospace industries. It urged the government to increase funding for research and technology "sufficient to maintain the UK industry's technological competitiveness".

The DTI has studiously avoided any financial commitment. It recently decided to stop funding industrial research directly through the Advanced Technology Programme, run jointly with the SERC. But Sainsbury's announcement indicates that

Volcano plans

By comparison, Japan's next five-year programme for volcanic eruption — to be announced the same day as the earthquake plan — seems almost radical. It designates regions for intense observation, such as the area around Mount Unzen which erupted in June 1991 and continues to erupt. It also calls for greater emphasis on basic research on magma chambers and the movement of magma, instead of collecting vast amounts of empirical data for prediction. **D.S.**

mittee for earthquake prediction, told the press that it might designate the region for special observation in the next five-year plan. This now seems unlikely.

David Swinbanks

the government has responded more positively to another recommendation of the select committee: that it should design a "single national technology acquisition plan" covering the technology needs of both civil and military sectors.

The government's statement is the first significant example of DTI's desire to take a more interventionist approach to industrial policy. In particular, Sainsbury expressed support for the principles behind a ten-year initiative, known as the National Strategic Technology Acquisition Plan, (NSTAP) drawn up by a committee of industry and academic advisers.

The plan identifies three categories of technology which, it says, UK industry will need over the next 20 years: fundamental technologies (such as advanced wing design) to provide competitive edge; "enhancing technologies" to improve the effectiveness of the industry; and supporting technologies imported from other sectors of industry.

According to officials from British Aerospace (BAe), the French government provides three times more research and development support to its domestic aerospace industry than Britain does, and Germany provides even more. The NSTAP, reflecting the views of companies such as BAe and Westland, urges the DTI to quadruple its own spending on research and technology acquisition in the field, from the current level of £22 million a year to about £100 million a year for the next ten years.

If SERC gets involved, it will probably be through its new Innovative Manufacturing Initiative. This is run under the auspices of a committee headed by Stewart Miller, director of engineering and technology at Rolls Royce. The SERC has already said aviation technology will be part of the IMI's core research programme. **David Dickson**

Moscow scraps Indian rocket deal under US pressure

Washington & New Delhi. A Russian decision, under US pressure, not to sell advanced rocket technology to India has provoked resentment in New Delhi, dismay in Moscow and contentment in Washington.

But the development has smoothed the way for collaboration between Russia and the United States on various space projects, although it is still uncertain what shape this cooperation will take.

Earlier this month, the United States and Russia settled the long-standing dispute over Russia's 1991 contract to sell to India cryogenic rocket engines and the technology to

ment when he visits the United States in August or September. On the same day, the Russian space agency and the National Aeronautics and Space Administration agreed to identify areas for cooperation. They will do so in time for Chernomyrdin's visit.

During the often heated negotiations on the Indian rocket agreement (whose value may be as much as \$400 million), Russian officials repeatedly demanded compensation for the revenue Russia would lose. Neither side admits to a *quid pro quo*, but the United States will now probably buy Russian space hardware.

Russia also hopes that the United States will put the station in an orbit that Russian rockets can launch to. They could then help build and/or eventually service it. Congressman George Brown (Democrat, California), chairman of the Science, Space and Technology Committee House of Representatives and an influential supporter of the space station, hinted last week that the United States could quickly agree to this.

But Roald Sagdeev, a former Soviet space official who heads the Washington-based East-West Science and Technology Center, warns that the US Congress or the international partners in the space station could still block major Russian involvement. His "greatest concern" is that US aerospace companies might see Russia as a competitor for "their piece of pie".

In Moscow, some see Russia's capitulation as betrayal. Nikolai Semyonov, an official of Glavkosmos, the commercial arm of the Russian space programme, was reported as saying that "they [the United States] are trying to tie a dead cat around our necks": one fear is that the United States is trying to stifle the Russian space industry to protect its markets.

In India, scientists at the Indian Space Research Organization (ISRO) understandably feel cheated. Russia had assured them, as recently as this spring, that the deal would go through. The rocket engines are critical to ISRO's plan to build a geostationary launch vehicle (GSLV). Russia will renegotiate the contract in New Delhi next month, but ISRO's chairman, Professor U. R. Rao, says that, without the technology, "the contract is as good as scrapped; there will be nothing further to negotiate."

Even so, Russia is said to have already delivered more than half the technical drawings for the rocket, and some in India believe it could develop the technology itself. "We have the capacity, but it will take time and much money," says R. Narasimha, director of India's National Aeronautical Laboratory in Bangalore.

Tony Reichhardt & K.S. Jayaraman

French state sells shares in Elf Aquitaine and Rhône-Poulenc

Paris. France's newly elected Conservative government has kept its pledge to speed up the sale of state holdings. First on the block are the chemicals and pharmaceuticals giant Rhône-Poulenc and the oil-to-chemicals company Elf Aquitaine — scheduled to be sold in September. Together they spent over FF10.6 billion on research and development last year.

Officially, both companies say the sale of state-owned shares will have little effect on either business or R&D strategies. But both have most of their research based in France, yet are pushing to internationalize their market influence; it would be surprising if privatization did not affect their R&D capacity in France.

Rhône-Poulenc — which was partly privatized earlier this year leaving the state with 43 per cent of the shares — invested FF5.96 billion in R&D last year. It has 19 major research centres worldwide, nine in France. But only four of its 11 life science centres are in France. "There will not be a step change but the globalization of R&D, which started in 1986 when the company hit the international acquisitions trail, will continue to evolve," says one London-based chemicals analyst.

Rhône-Poulenc's R&D strategy involves tight links with French public research institutions. More than 300 researchers take part in the FF1.6 billion BioAvenir programme. BioAvenir is designed to develop quickly fundamental biology research of commercial promise. The company's privileged access to innovations from some of the world's most productive laboratories will not harm the value of the government's holding when it sells.

When BioAvenir was launched, critics alleged it was a guise for state aid. The company has since invited others to invest, but none has yet done so. Rhône-Poulenc has hinted it might create a similar programme to develop chemical R&D. But the French government may be less willing to invest in a company in which it has no stake.

Elf-Aquitaine employs nearly 6,000 researchers from a budget of FF4.62 billion in 1992. It spends most — almost half — on health-care research, followed by chemicals (29 per cent) and hydrocarbons (18 per cent).

As a concession to President François Mitterrand, the French government will probably keep a golden share in the strategic group; it currently holds a half stake. Privatization, says a company spokesman, would probably not affect its business or research strategies. But the company supports fundamental research and has 520 research contracts with public research organizations.

Mike Ward

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Novosti Press Agency/Science Photo Library

The Russian Proton rocket could launch parts of the US space station.

build them. The United States argued that the sale violated international non-proliferation agreements because the engine technology could be used for missile launchers. India and Russia have dismissed the charge as absurd.

Nevertheless, Russia has now agreed to sell only assembled engines to India and not the technology. More important from the US standpoint, Russia has agreed to abide by the terms of the Missile Technology Control Regime (MTCR). The United States had feared that cash-hungry Russia might sell military technology indiscriminately: it was reported to have sold submarines to Iran and ingredients of rocket fuel to Libya.

The United States told the Russians that, unless they settled, it would withdraw an offer to allow them limited entry to the commercial satellite launch market. It also threatened to ban all US dealings with KB Salyut, the Russian partner in the Indian rocket deal. Because KB Salyut also designs space vehicle hardware, the ban would have blocked Russian participation in the US space station and shuttle programmes.

The day after Russia agreed to scrap its deal with India, the US unblocked the satellite launch agreement (see *Nature* 363, 661; 1993). The Russian Prime Minister, Viktor Chernomyrdin, is expected to sign this agree-

Use of fetal eggs in research to be debated

London. Scientists working in embryo research and infertility treatment could soon use eggs taken from aborted fetuses. In response to this development, the UK Human Fertilization and Embryology Authority (HFEA) last week said that it would ask the public to suggest how it should regulate the use of eggs from aborted fetuses in embryo research and infertility treatment.

Animal research suggests that the use of such eggs will soon be feasible for both applications. Similarly, doctors may soon be able to transplant fetal ovaries, potentially a radical new way to treat infertility. But those carrying out research into both techniques recognize that each raises new ethical issues.

Robert Edwards — the British pioneer of *in vitro* fertilization (IVF) — says he raised the idea of using fetal eggs for research in a paper published in 1975. Fertility laboratories have shown increasing interest in recent years because of shortages of eggs provided by donors for research.

"Fetal eggs have the potential for many medical and scientific uses," says Colin Campbell, professor of law at the University of Birmingham and chairman of the HFEA. "The authority's formal interest lies in their potential for use in embryo research and infertility treatment; however, profound ethical questions first need to be addressed."

Doing research with fetal eggs raises even more complex ethical issues than with other fetal tissue (in particular as a potential treatment for Parkinson's disease). A child, for example, could know its grandmother although its biological mother had never been born. Would the man who helped to create the fetus have any rights over his germ-line tissue?

The committee set up to establish guidelines for the use of fetal tissue in research under the physicist and theologian John Polkinghorne, who is president of Queens' College, Cambridge, did not consider the question of eggs. Some of its conclusions would almost certainly be inappropriate: for example that mothers should not be given details of how tissue from her fetus is subsequently used.

HFEA officials say they will set out the scientific prospects in a consultation document later this year; this will cover both fetal eggs and those taken from dead women. HFEA will use the responses in reaching its conclusions on how to regulate research and eventual treatment.

Roger Gosden, a senior lecturer in physiology at the University of Edinburgh who is pioneering research into the possibilities of ovarian transplantation, welcomes the HFEA initiative. Given the demand for infertility treatment, he says, the HFEA needed to clarify what sort of research and treatment society considers acceptable: "We need an

authoritative consensus view to clear the way, and to provide an ethical basis on which we can move forward."

Edwards also says that he welcomes the debate about using fetal eggs for research. He warns that scientists who proceed faster than public opinion is prepared to accept run the risk of creating a backlash. "Personally I am against working with fertilized eggs originating from fetal ovaries just yet," says Edwards. "I would like to move more slowly, for example using such eggs to study how

follicles develop, and that sort of thing."

In January, the HFEA consulted the public about sex selection techniques. Two-thirds of the respondents said these should be permitted for medical but not for purely social reasons. They thought it acceptable, for example, if a woman risked bearing a child with a life-threatening disease. HFEA has now incorporated this conclusion into its code of practice, which all clinics registered to perform *in vitro* fertilization must follow.

David Dickson

Australia reconsiders research plans

Sydney. The Australian government last week promised to reconsider the plans of the science minister, Chris Schacht, to reorganize public research, following protests from the scientific community and other cabinet members.

Schacht's plans include moving the Fisheries, Oceanography and Atmospheric Research divisions of the Commonwealth Scientific Industrial Research Organisation (CSIRO) into the Australian Institute of Marine Science (AIMS). By creating a giant marine research institute, he hoped to boost marine and fisheries research, and increase Australian exports.

He also wanted to make the Australian Nuclear Science and Technology Organisation (ANSTO) part of CSIRO. This he believed would improve ANSTO's poor public image: but this has more to do with public loathing of nuclear technology than anything else.

But the Federation of Australian Science and Technology Societies and the Public Sector Union complained that the government had not properly consulted the organizations affected; these heard of the plans only a few weeks before a government committee was to consider them. Schacht says he did not make his plans public because this would have given CSIRO more opportunity to attack them. There have been too many reviews of Australian science, he says.

Other politicians supported the scientists' cause: Senator Gareth Evans, the Minister of Foreign Affairs, opposed the CSIRO's takeover of ANSTO, while the Federal Minister for the Environment, Ros Kelly, attacked a plan to take the Division of Environmental Research out of CSIRO. The media accused Schacht of trying to make his mark in his new job.

The federal government committee has now referred Schacht's proposals to an internal and a public review. Professor Ken McKinnon, vice-chancellor of Wollongong University, will examine merging CSIRO's marine divisions with AIMS. He will also consider an alternative proposal to incorpo-

rate AIMS into CSIRO. The government will start its internal review of ANSTO after it resolves the separate issue of whether to give the organization a new nuclear reactor.

Despite this setback, Schacht denies that the government has rejected his plans. He says he welcomes the exchange of views they have prompted.

Mark Lawson

Malpractice denied

Paris. The Pasteur Institute in Paris has denied any malpractice over the deaths from Creutzfeldt-Jakob disease of several children who were treated with human growth hormone between the end of 1983 and the middle of 1985. This followed the French government's decision last week to charge the two chief doctors responsible for the production and distribution of the growth hormone with involuntary homicide over the deaths of three children who were being treated for dwarfism.

One of those charged is Jean-Claude Job, president of the organization France-Hypophyse (France Pituitary), which was responsible for collecting pituitary glands from cadavers in France and elsewhere, and distributing the hormone extracted from them. The other is Fernand Dray, head of the laboratory at the Pasteur Institute where the extraction and purification was carried out.

Last December, the Ministry of Health ordered an inquiry following reports that some of those treated with the hormone died from the nervous disease eight years later. Its report strongly criticized the institute for the conditions under which it processed the pituitary glands.

But the institute claims that the onset of the disease in 24 patients treated with the hormone — 19 have since died — was "the consequence of the risk attached to a new therapy". It says the nature of the organism causing Creutzfeldt-Jakob disease was (and is) unknown, and so it was impossible to screen the batches of hormone to discover if they were contaminated.

David Dickson

US Congress urged to back further Agent Orange studies

Washington. A report prepared for the US Congress by the Institute of Medicine (IoM) this week called for yet further investigation of the effects of the defoliant Agent Orange on the health of Vietnam veterans.

Congress commissioned the 500-page report in a bid to clear up the long-running controversy over Agent Orange, a herbicide the Americans sprayed extensively during the Vietnam war. The report overturns previous recommendations by several official agencies that further investigation would be futile because it was difficult to gather reliable data on exposure of veterans.

The US Department of Veterans Affairs already compensates exposed veterans who have developed non-Hodgkin's lymphoma, soft tissue sarcoma, and chloracne. But the report says there is now sufficient evidence to link exposure with Hodgkin's disease and the liver disorder porphyria cutanea tarda (PCT) as well.

But for a far longer list of alleged health problems — including various cancers, infertility and birth defects — it says there is insufficient evidence to establish such a link.

House and Senate veterans affairs committees will this week start hearings to consider the report. The Department of Veterans Affairs is expected to respond within 90 days, possibly by widening its compensation rules.

An IoM panel prepared the report for the National Academy of Sciences. The inquiry was intended to inject some balance into what historically has been a highly polarized debate between angry veterans and sceptical government agencies. Veterans' organizations were furious, for example, when the Centers for Disease Control concluded in 1989 that it was impossible to untangle the Agent Orange episode (see *Nature* 340, 179; 1989).

The panel, chaired by Dr Harold Fallon of the University of Alabama, recommends that a nongovernmental body should look for new ways to estimate exposure of veterans, particularly those not involved in spraying. An independent panel should then evaluate these techniques, it says; if they found them adequate, the government should then start an extensive epidemiological study to assess possible links between Agent Orange and cancer and other diseases.

Dr David Tollerud of the University of Pittsburgh, vice-chairman of the panel, says that such an investigation, deemed impractical a few years ago, is now worthwhile: "Several things have changed. First, we got information that substantial amounts of herbicide were used around US bases and on

waterways, but not recorded. Second, newer approaches in epidemiology allow us to combine informal as well as formal data in models. Finally, we have better computer hardware and software. So it may be possible to develop a useful model."

"We are not saying that a big epidemiological study should be done," says Tollerud. "We just outline various steps that should be taken in preparation for one." The IoM panel did not cost the preparatory work or the study, or suggest their priority compared

with other health research.

Congress asked the National Academy of Sciences only to review and evaluate existing scientific and medical information on Agent Orange. But the panel made research recommendations because, Tollerud says, "we felt an ethical obligation to go beyond just stating the scientific facts".

The report touches briefly on prospects for research on the effects of Agent Orange on the health of Vietnamese civilians: the United States sprayed 11.2 million gallons on their homes and crops between 1962 and 1970. But Tollerud says the report did not recommend a Vietnamese study because an investigation of US war veterans "is more tractable and more relevant to what we were asked for".

Colin Macilwain

Russian science nearing collapse

Moscow. Russian science may now be on the brink of collapse, following the suspension of payments by the Russian Academy of Sciences (RAS) to its dependent institutes.

In many institutes, salaries are not now being paid. At the Institute for Molecular Genetics in Moscow, for example, employees have been given extended leave without pay over the summer months. At the Institute of Applied Mathematics, the administration is paying only 40 per cent of agreed salaries. Elsewhere, funds intended for equipment and supplies are being spent on salaries instead.

Both the Minister for Science, Boris Saltykov, and the president of the RAS, Yuri Ossipov, have used the phrase "brink of collapse" in their comments on this situation, which has arisen unexpectedly, and after a short spell when Russian scientists sensed some improvement of their financial condition for the first time since the reforms began.

What has gone wrong? At the beginning of the year, the budget of the RAS was increased nearly threefold, from 250 billion rubles to 680 billion rubles. Although this was not agreed by the Supreme Soviet, payments by the academy in the first half of the year were based on the increased budget. Moreover, by a decree of President Boris Yeltsin, the salaries of all RAS employees were increased by a factor of 1.9.

Now, it seems, these promises will not be kept. At the end of June, the finance ministry told Saltykov's office that the index factor would be 1.47, not 1.9. In the meantime, the RAS ran out of funds. Institute salaries for May were not paid. Only early in July was there a payment to cover salaries for May and part of June.

The effects of the payments freeze vary from one institute to another. Those earning a large proportion of their income from sources other than the RAS may be able to

weather the crisis without serious disruption. But reports from other institutes describe hardship and even hunger among researchers.

Ossipov has told the newspaper *Izvestiya* that, in this precarious situation, the Russian academy is "balanced on the brink of unstable equilibrium, where a slight push will cause everything to collapse". Institute researchers feel badly let down by the government. The research community gave Yeltsin its support in the hard times before the referendum in April, and had believed that better treatment for the research community would be their reward.

Now the mood among researchers has shifted. At an extraordinary conference on 30 June, the RAS trade union called on its members to be ready to strike, and demanded of the government and of Yeltsin that an effort should be made to settle financial relations between the state and science. An unprecedented protest rally is being arranged for 14 August.

Alienation of the research community could have serious consequences for the Russian government. Although the scientific intelligentsia has hitherto been one of the chief sources of support for democratic reform, there are now ominous signs that some of its members are moving towards what Yeltsin calls the "irreconcilable opposition".

Although the views of the union leaders at the meeting on 30 June do not represent those of the scientific community, the expressions of nationalism heard at the meeting are worrying. One opinion was that Russian science is being deliberately starved in order to crush the Russian nation "by the castration of its brains". Many will interpret these sentiments as a measure of people's desperation, but if they are widely echoed, they could become dangerous.

Vladimir Pokrovsky

Key statistician ousted as Waxman enters gp160 fray

A statistician involved in Army AIDS research work has resigned, alleging that his employer, the Jackson Foundation, is party to an Army "cover-up" of information surrounding the release of early data on the efficacy of the gp160 therapeutic vaccine.

Bill McCarthy, director of biostatistics at the Maryland-based foundation — a nonprofit, 600-strong clinical research centre which works primarily for the armed forces — says that it has failed to confront what he regards as shortcomings in data presented by Lt Colonel Bob Redfield, the researcher at the centre of the continuing gp160 controversy, because of its reliance on Army money. McCarthy has been an internal critic of Redfield's work for some time, and his resignation from the Jackson Foundation appears to have been forced after a row there over the alleged leaking of information to the press.

Meanwhile, Congressman Henry Waxman (Democrat, California), chair of the House health and environment subcommittee, has written to health secretary Donna Shalala expressing his concern about the fate of a proposed large-scale gp160 vaccine trial, and asking for an update on it. Responsibility for the \$20 million trial, which was transferred earlier this year from the Army to the National Institutes of Health, has now reverted to the Department of Defense (see *Nature* 363, 294; 1993).

McCarthy's exit appears to mark a victory for the Army research faction that endorses Redfield's belief in the efficacy of gp160 vaccine as a treatment for HIV-infected patients.

But after McCarthy resigned, he made public a series of allegations about the management of Army AIDS research and the role of his former employer. McCarthy says that:

- Data presented by Redfield at the international AIDS conference in Amsterdam last year was at odds with the actual results of the Army's Phase 1 gp160 trial;
- The subsequent Army investigation, whose results have yet to be officially released but which is known to have exonerated Redfield of scientific misconduct, was a "whitewash" which failed to adequately address the legitimacy of the Amsterdam paper;
- Management at the Jackson Foundation, where Redfield and his Army colleagues conduct much of their AIDS research, has turned a blind eye to this because of its reliance on Army funding.

McCarthy says that his own statistical analysis of the results of the Phase 1 gp160 trial shows that "done properly, you could not get the results given by Redfield in Amsterdam or subsequently".

The army investigators, he claims, "gave

a sense that the investigation was primarily a damage control exercise." Of John Lowe, the Jackson Foundation's director, he says: "Whenever I tried to get him to do the right thing, he'd get mad at me for jeopardizing [the Foundation's] Army connection."

Lowe has not responded to requests for interview. Lisa Reilly, a spokeswoman for the Jackson Foundation, says: "Dr McCarthy resigned for personal reasons. The foundation has a policy of not discussing matters regarding former employees".

Discounts offered for large-scale users of Taq

Washington. Roche Molecular Systems (RMS) and Perkin-Elmer Corporation this week agreed to offer discount on *Taq* DNA polymerase for large-scale users, because of pressure from the scientific community. The companies also hope that cutting the cost of the enzyme, which drives the polymerase chain reaction (PCR), will stimulate new applications for PCR.

Under the bulk discount programme, Roche and Perkin-Elmer will drop the unit price of AmpliTaq *Taq* polymerase for laboratories that purchase more than 250,000 units of the enzyme a year; above this, the price per unit will be further discounted depending on annual consumption.

Through a system of matching volume-for-volume grants, the companies plan to cut the unit price of the enzyme by half again to all designated National Center for Human Genome Research (NCHGR) Genome, Science and Technology Centers (see figure for the new pricing schedule).

Stanley Rose, director of PCR products in the Applied Biosystems division of Perkin-Elmer does not anticipate any sig-

A spokesman for the Walter Reed Army Institute, Marvin Rogul, denies there has been any cover-up, and defends the Army record on AIDS research. "When you have clinical trials there is always some room for differences of opinion," he says. Referring to Redfield's release of efficacy data at Amsterdam, which McCarthy says was premature, Rogul says: "If you don't give out data you get accused of holding it back."

Rogul says that the Army was still negotiating with gp160 producer MicroGeneSys about the provision of free vaccine for a single product, Phase 3 trial. If negotiations fail, he points out, the \$20 million allocated for the trial will revert to the Army AIDS research budget. Comment on how that money might be spent would be "presumptuous", Rogul says. **Colin MacIlwain**

nificant delays in implementing similar pricing schemes for large-scale genome mapping and sequencing laboratories in Europe and Japan.

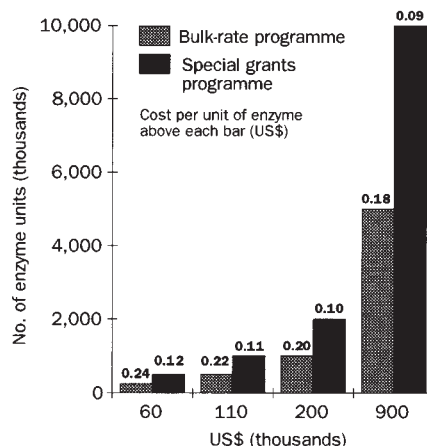
In a related development last week, the companies loosened their stranglehold on the supply of *Taq* polymerase. They granted Boehringer Mannheim a non-exclusive licence, valid for the term of Roche's PCR patents, which provides the company with worldwide rights to make and sell the enzyme for use in PCR in all fields other than *in vitro* diagnostics.

Boehringer Mannheim already sells *Taq* polymerase for non-PCR applications (although it cannot police how researchers use it). Now it can compete with Perkin-Elmer which previously had sole rights to sell PCR products made by RMS in these markets.

Researchers greeted news of the pricing programme enthusiastically. Dr David Schlessinger of the department of molecular microbiology at Washington University School of Medicine called it a "salutary development". Dr Francis Collins, who in April succeeded James Watson as director of the National Institutes of Health's NCHGR, called it a "win-win" situation. In a recent poll of 13 genome centres, which were paying an average of \$0.30 per unit of enzyme, Collins estimated the centres were spending as much as 12 per cent of their budget on *Taq* alone; under the new pricing scheme, the cost per unit for a centre using, for example, more than one million units annually would drop by about a third to \$0.10.

President of the Human Genome Organisation and director of the institute for molecular genetics at Baylor College of Medicine, Dr C. Thomas Caskey, says he had hoped Roche would open up the market to more than one other supplier. Nevertheless, "it's a step forward," he says, acknowledging that RMS has come a long way over the past year or so: "I think they realize that there were some mistakes made early on", he says. **Diane Gershon**

Pricing scheme for Taq



Caribbean hot spot

SIR — There are no natural forests remaining in Haiti, apart from a few isolated patches, and the total forest cover is below one per cent^{1,2}, making Haiti's forests among the most severely depleted ecosystems on Earth. By comparison, forest cover in tropical Central and South America averages 27 and 44 per cent respectively². Haiti's neighbour on Hispaniola, the Dominican Republic, is in slightly better ecological shape, but many Haitian species are found only in Haiti³, and mass extinctions of plants and animals are likely to occur in the next few years in Haiti unless measures are taken.

The growing human population of Haiti (7 million), and its demands on resources, is the long-term problem that must be addressed. But the immediate concern is the felling of trees, primarily for the production of cooking fuel (charcoal). Alternative sources of energy need to be found if the little remaining forest is not to be exhausted. Many species still exist in these small areas of remnant forest^{3,4} and many can be saved if there is prompt action to slow or halt the deforestation. We propose the following action:

(1) Funding for biological conservation by the US Agency for International Development (USAID) in Haiti, cancelled as a result of the *coup d'état* and subsequent OAS embargo, should be resumed and increased. Priority should be given to preserving biological diversity in the remaining pockets of forest cover, mostly

located in or near the existing national parks, and finding alternative sources of energy for the local inhabitants.

(2) The World Bank Forestry and Environmental Project, suspended at the time of the *coup*, should be reinstated. This project was designed to upgrade the Service for the Protection of the Environment in Haiti, rebuild the infrastructure of the Ministry of Agriculture associated with the protection of biodiversity and initiate a training programme in natural resources and wildlife.

(3) The University of Les Cayes, near the last remaining areas of natural forest, should develop a curriculum in the stewardship of natural areas. The national parks at Pic Macaya and Morne La Visite could be the 'living laboratories'.

(4) A private, non-governmental organization should be established to serve as a clearing house for information on natural history and conservation, thus ensuring that this information is available to other similar organizations, international donor programmes and the government of Haiti.

(5) A strong programme in environmental education should be established.

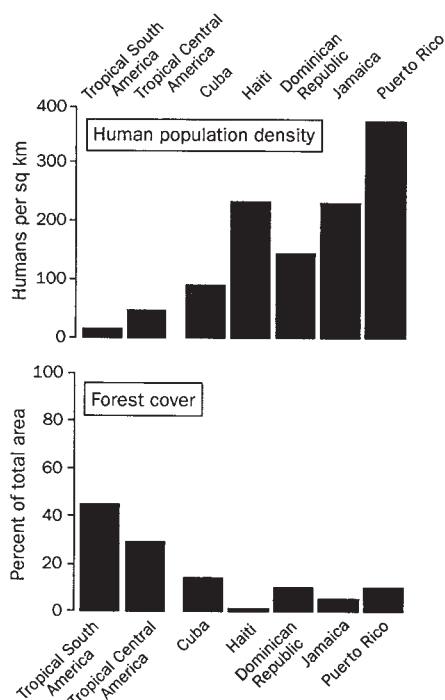
If these steps are taken immediately, the pending mass extinctions in Haiti may be delayed or even prevented. The infrastructure is in place, and studies to establish the programmes have been carried out⁵. International donor agencies such as USAID must resume their support for environmental programmes in Haiti. Even if the political situation remains confused and controversial, four of these five programmes could be implemented. Only the World Bank Project depends on negotiations with a stable, internationally recognized government of Haiti.

S. Blair Hedges

Charles A. Woods

Department of Biology,
208 Mueller Laboratory,
Pennsylvania State University,
University Park, Pennsylvania 16802, USA
and
Florida Museum of Natural History,
University of Florida,
Gainesville, Florida 32611, USA

1. Paryski, P., Woods, C. A. & Sergile, F. in *Biogeography of the West Indies: Past, Present, and Future* (ed. Woods, C. A.) 855–878 (Sandhill Crane, Gainesville, Florida, 1989).
2. World Resources Institute. *World Resources 1992–93* (Oxford University Press, New York, 1992).
3. Woods, C. A. & Ottenwalder, J. A. *The Natural History of Southern Haiti* (Special Publ. Florida Mus. Nat. Hist., Gainesville, Florida, 1992).
4. Hedges, S. B. *Biodiversity and Conservation* 2 (in the press).
5. Woods, C. A., Sergile, F. & Ottenwalder, J. A. *Stewardship Plan for the National Parks and Natural Areas of Haiti* (Florida Museum of Natural History, Univ. Florida, Gainesville, Florida, 1992).
6. Caribbean/Central American Action. *Caribbean and Central American Databook* (Caribbean/Central American Action, Washington, DC, 1990).
7. Johnson, T. K. *Biodiversity and Conservation in the Caribbean: Profiles of Selected Islands* (International Council For Bird Preservation, Cambridge, 1988).



Recent estimates of human population density and forest cover in the Greater Antilles, with comparable estimates in Central and South America^{1,2,6,7}.

Grant policy

SIR — Your article (*Nature* 363, 4; 1993) "Wolfson Foundation's policy on animals angers other British medical charities" stated that the Physiological Society "would accept the foundation's argument" in not supporting intercalating medical and dental students who do research projects on experimental animals. We have decided for the time being not to pursue the matter, but we do not agree with the Wolfson's policy.

The society believes that the use of experimental animals is vital for progress in medical research. It also believes that the intercalated year is an important first step in the training of doctors and dentists for research. The trustees of the Wolfson Foundation plainly have the responsibility for setting their own policy. Over the past few years, they have taken an important and welcome lead in supporting intercalating studentships. However, we believe that their policy on projects involving experimental animals is regrettable and mistaken.

G. J. Dockray

D. A. Elsner

D. Cotterrell

The Physiological Society,
PO Box 506, Oxford OX1 3XE, UK

Minister's boson

SIR — So Britain's cabinet minister for science wants to hear from scientists who can explain on a single sheet what the Higgs boson is (*Nature* 362, 781; 1993)! No doubt this herculean challenge will keep British physicists busy for some time. For what chance of success do they have in attempting such simplification of science?

And why do politicians on this side of the Atlantic nearly always demand that scientists demonstrate the value of scientific research year after year — this is really what they mean by simplification and popularization — when we ask for continued financial support? Many cogent studies and arguments have shown the economic and social benefits of scientific research, including *Retrospectroscope: Insights into Medical Discovery* by J. H. Comroe Jr (Von Gehr Press, 1977). This book still offers an excellent insight not only into "how to get the most for the medical research dollar" but also for the general research pound, franc, lira and so on. Regrettably, I am unable to recommend the book to the Irish minister of science; the office was abolished following the recent general election; it is now technology only and our science grant was cut by more than IR£4 million in the February budget.

James J. A. Heffron

Department of Biochemistry,
University College Cork,
Cork, Ireland

Know your Japanese host

SIR — Your feature on “Employment in Japan” (*Nature* **362**, 867–870; 1993) did a good job in describing the opportunities available for foreign scientists to work in Japan. To be complete, however, your article should have mentioned the experiences of previous fellows so that applicants would know about potential difficulties. Having spent more than two years at a Japanese national institute, first as an AIST (Agency of Industrial Science and Technology) fellow and now as an STA (Science and Technology Agency) fellow, I’ve been able to see and hear about many commonly encountered problems.

You warn prospective fellows that “applicants should take care to avoid institutes best described as ‘dinosaurs’.” A far more important consideration is choosing the Japanese host researcher, who is the individual with whom the researcher most often comes in contact and who is responsible for the working conditions. A sour relationship with a host guarantees a miserable fellowship. Instances of Japanese hosts using their foreign researchers as laboratory technicians, or ignoring them completely, while spending the fellow’s research funds, although certainly in the minority, are not uncommon. More serious situations have occurred in which the host insists on exerting undue control over a fellow’s lifestyle both inside and outside the workplace. To be fair, such problems are not unique to Japan. But I would strongly advise a potential applicant to have made some face-to-face contact before making a final decision. Finding a compatible host is, I believe, the single most important factor in determining a successful fellowship.

Another complaint is the geographical isolation of Japan. Most visiting researchers come as postdoctoral fellows and this isolation acts as a barrier to finding a permanent position elsewhere. Foreign journals containing job announcements sometimes arrive after the deadline for applications. Travelling to an interview or to attend a conference is often prohibitively expensive. It is easy for a foreigner in Japan to feel out of the mainstream and to lose contact with colleagues and potential employers elsewhere.

Finally, the language barrier is formidable. It is rare that an individual arrives already fluent in Japanese. Most fellows arrive with little knowledge of Japan let alone the Japanese language. By relying only on English, the potential for misunderstandings is great. Patience and flexibility on the part of the researcher, as well as a willingness to learn the language and customs of Japan, are required. In addition, the language barrier creates an even more serious problem for the researcher’s family. Opportunities for employment or

socializing are limited at best and the spouse frequently complains of boredom. Even outside the workplace, the language barrier affects the researcher.

None of this should dissuade prospective fellows from coming. I believe most fellows (myself included) have found their stays to be professionally rewarding. But in order to prevent disappointment or hard feelings between future fellows and their Japanese colleagues, it is essential that foreign researchers contemplating a long-term visit to Japan should be aware of the problems that can arise.

Philip A. Jarvis
*Geological Survey of Japan,
Higashi 1-1-3, Tsukuba,
Ibaraki 305, Japan*

Joking apart

SIR — US readers of *Nature* rightly considered it scandalous when *Nature* recently stated (**363**, 480; 1993) that “the US public rightly considered it scandalous that their vice president could not spell potato and thought that Latin Americans speak Latin. . .” because, while the first really did happen, the second was only a joke by Johnny Carson on the Tonight show.

Only a scientist could confuse reality with television in a leading article about the need for factual accuracy.

Michael Broyde
*Department of Religion,
Emory University,
Atlanta, Georgia 30322, USA*

UNESCO

SIR — Your leading article (*Nature* **363**, 382; 1993) on recent steps taken by the United States to rejoin UNESCO was disappointing. Both the United States and Britain have left the organization in protest against Dr A. Amadew-Mahtar M’Bow’s policies. It would seem though, that his replacement by the distinguished Spanish biochemist Dr Federico Mayor has so far failed, despite many years of trying, to persuade the two countries to rejoin.

UNESCO’s contributions to the developing countries can be largely felt and experienced by people living in these countries. I was made aware of some of UNESCO’s contributions to the Middle East while holding an academic senior position in which I was responsible for research and development. Its advice, help and contribution to many cultural and scientific ventures was plain to many in the Middle East, ranging from attempts to restore and preserve ancient towns and cities to the organization of visits to places

such the International Centre for Theoretical Physics in Trieste. In fact, UNESCO’s contribution seemed to be complementary to the greatly appreciated works of organizations such as the British Council and the US Agency for International Development.

The United States and Britain should reconsider their position and rejoin UNESCO as soon as possible. I am sure that their contribution, at all levels, would be greatly valued. Finally it is evident that the exact changes sought by both countries can be met only by working from within and not by simply staying out.

M. S. Suleiman
*University of Bristol,
Department of Physiology,
Park Row, Bristol BS1 5LS, UK*

Native rights

SIR — I was annoyed to read (*Nature* **363**, 388; 1993) that the government of Queensland (Australia) is considering legislation declaring “its right to a share of any financial gain made from research involving native plants and animals”. Surely, the rights to native plants and animals belong to the *natives*?

Native people around the world have been the victims of European aggression and expansionism for centuries. The looming battle over “ownership” of native species (or more correctly, their products) could be an opportunity for the world to recognize (finally) native rights to native materials. The English-speaking invaders of Australia and North America as well as the Spanish- and Portuguese-speaking invaders of “Latin” America have stolen the land rights (mining, logging and even water) of the people who most civilized thinkers should admit are the rightful owners.

Today, once again, there are rumours of great riches in the jungles and the Biotech Conquistadors will walk all over the rights of natives in order to steal their treasures. They will enlist the knowledge of the few remaining native wise-men for clues as to exactly which species contain the prize. The invaders, well versed in the weaponry of public relations, will argue that some profits will trickle down to the poor natives or promise that the profits will be used to create nature reserves (designed, administered and enjoyed mostly by European invaders). The rights of natives will again be abused in order to enhance the wealth of the conquerors.

Jamie Love
*76 Main Street,
Roslin, Midlothian EH25 9LS, UK*

Letters submitted for Correspondence should be typed, double-spaced, on one side of the paper only.

What road ahead for Korean science and technology?

South Korea's rise to the ranks of industrialized nations, symbolized by next week's opening of a science and technology exposition in Taejon, has been meteoric. But the road ahead is fraught with difficult choices. Many policy makers want the country to continue to focus on short-term development of technology to sell on the world market, but there is also a clear need to build a broader base for more basic long-term research.

YOU can measure the pace of development in South Korea by the length of the traffic jams in Seoul and other major cities. Four decades ago, at the end of the Korean war, the country was a poor undeveloped agricultural society with no industrial base. Now, bumper-to-bumper lines of Korean-made Hyundai cars fill the 8-lane roads of central Seoul (despite a well-developed system of subway trains), reflecting the devotion of South Koreans to technology.

But rubbing shoulders with giants brings problems of its own. As the United States, Europe and Japan are clamping down on the transfer of technology to South Korea and demanding better protection of intellectual property rights, it is becoming more difficult and costly for companies to license foreign technology. It is not therefore surprising that government policy makers and industrialists are calling for greater investment in science and technology, with the emphasis on technology. And the rising wages that have followed economic success and the rapid democratization of the country in recent years have made South Korean products less competitive on world markets, which is also driving industry to develop value-added technology.

Echoing previous presidents, the new administration of President Kim Young-

sam, who took office earlier this year, is calling for government-industry investment in science and technology to be doubled (in percentage terms) to 5 per cent of gross national product (GNP), or about \$42 billion a year, by 2001.



History suggests that this ambitious goal, which would match total German investment in R&D in 1989, can be achieved, particularly if President Kim, who enjoys enormous popular support, puts his weight behind the policy. In 1981, total investment in science and technology by the government and private sector was only about \$0.7 billion a year, or under 1 per cent of GNP, but

now it stands at \$7.6 billion a year, or 2.6 per cent of GNP, according to statistics of the Ministry of Science and Technology (MOST).

Much of the growth has been in the private sector, which now accounts for over 80 per cent of research and development spending, up from about 60 per cent ten years ago. Science advisors to Kim want to reverse this trend and increase the government share. Kim, the first president not drawn from the powerful military, has devoted the first months of his administration to rooting out chronic corruption in government, industry and the military, and he has surprised all by his ability to do so. "I can't understand where he draws his power from", comments one research leader in industry, "ten years ago this country was a military dictatorship."

The president has said that science and technology policy will be a top priority of his administration. A major restructuring of the government's management of science and technology appears imminent and this is expected to provide an opportunity for a boost in government spending in this sector.

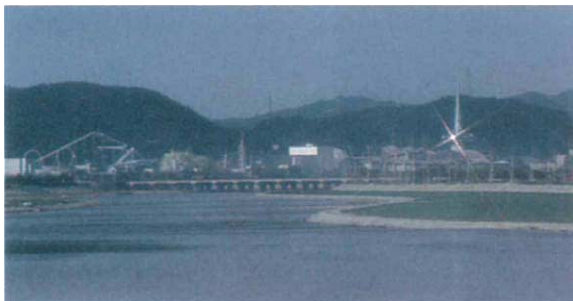
The prospects for Korean science and technology are thus bright. But crucial decisions lie ahead about the ways in which the proposed extra money should be spent and administered. While some advocate targeting the development of specific high technology to catch up with advanced Western nations and Japan, others think such matters should be left to South Korea's already successful and powerful industry. They instead point to the dire need for the government to reform the higher education system, particularly at the graduate school level, and to invest more in basic long-term research.

South Korea is often compared to its powerful neighbour Japan and is said to be about ten years behind. While it is true that some of the fashion of women in Seoul is reminiscent of Tokyo ten years ago, there are many areas, for example, megabit semiconductor memory chips, cars and steel, where the two countries are competing neck and neck.

Indeed, South Korea's development has been far more rapid and dynamic than its neighbour's. While Japan has had

Taejon Expo'93 — ambition on display

The first international exposition of science and technology to be held in a newly-industrialized country opens next week (7 August) in Taejon in central South Korea. Over 100 countries and numerous international organizations will participate in the exposition which has the theme "the challenge of the new road to development" and two sub-themes "traditional and modern science and technology for the developing world" and "towards an improved use and recycling of resources".



Taejon's Expo — In the limelight

The South Korean government has invested over \$500 million in the event, including a new highway linking the site with Seoul, which is expected to draw 10 million visitors during its 3-month period of operation. The site is located next to Taeduck science town which has many new government and private sector research institutes and will be absorbed into the town after the event. The exposition will feature South Korea's Highly Advanced National Project, otherwise known as the G-7 project, which aims at catching up with the G-7 nations (Japan, United States, Canada, United Kingdom, France, Italy, and Germany) in certain technologies by 2001. Advanced Western nations exhibiting at the site will focus on sustainable development and Earth surveillance to protect the global environment.

125 years to reach its present status, South Korea has taken a mere thirty years to become a strong world competitor in such industries as steel, automobiles, shipbuilding, and electronics. Growth of the infrastructure for research and development has been equally rapid.

The pace of administrative change is often bewildering. South Korean government research institutes are lucky if they manage to retain the same name for more than a few years without being split up, merged or put under the jurisdiction of another ministry. The handful of people who initiated this dynamic process in the 1960s, many of them educated for several years in the United States and/or Japan, still hold positions of power in government and industry and are involved in formulating future policy.

This contrasts with Japan, which in some ways is still emerging from the self-imposed isolation of the 200-year Tokugawa era and where government officials are only now seeing the need to internationalize their science and science policy-making activities. (Although Japan did flirt briefly with internationalization during the Meiji restoration last century when it absorbed Western science and engineering).

For a few years after the Korean war (which ended in 1953), the South Korean government tried to make progress by developing agriculture. But the country's leaders soon realized that the only way out of poverty was to develop a strong infrastructure of industry based on science and technology. South Korea was one of the first countries to join the International Atomic Energy Agency in 1957 and the government's first research institute, the Korea Atomic Energy Research Institute, was established two years later. (Currently, nine nuclear power plants provide nearly half of the country's electricity).

The first big push to establish an infrastructure for government research came in the mid-1960s, under the initiative of then President Chung Hee Park, with the support of the United States. In return for help in the Vietnam war, US President Lyndon Johnson in 1965 sent a team of US scientists led by the president's science and technology advisor, Donald Horning, to help set up the Korea Institute of Science and Technology (KIST). A detailed plan was developed by Battelle Memorial Institute of the United States. KIST opened in 1966, and has since spawned a network of affiliated institutes with close to one thousand researchers and engineers, as

well as several independent research institutes.

In parallel with the establishment of KIST, the Ministry of Science and Technology (MOST) was created in 1967 from a bureau in the powerful Economic Planning Board to manage and coordinate government research. MOST has played a central role in the country's science and technology policy, but has always been comparatively weak compared with larger existing ministries, such as those of trade and industry and communications.

KIST's initial role was to select and digest good technology to support South Korea's fledgling industry, says Hyung-Sup Choi, the first president of KIST and a former minister of science and technology (1971–1978). One early project was to design a cheap 200-ton boat to be made of concrete. But by the mid-1970s, KIST was being commissioned to design things like 250,000-ton oil tankers for the giant Hyundai industrial group, and so several new specialist institutes, for shipbuilding, oceanography, telecommunications technology, mechanical engineering, chemical research and standards research were spun off.

By the late 1970s, KIST became almost viable, with about 80 per cent of its funds

KIST – what future role?

THE Korean Institute of Science and Technology (KIST) is the "mother of technology development in Korea" says Hyung Sup Choi, first president of the institute and a former minister of science and technology. But now KIST must find a role in a "higher dimension", he says.

After a bright start in the 1960s and 1970s backed by South Korea's then president Chung Hee Park, KIST seems to have lost its bearings a bit in the past decade. The unsuccessful merger with KAIST (see above) did not help, nor did the government's decision to pump money into the institute in the early 1980s when the institute was almost self-sufficient. Now KIST has to depend on the government for about 80 per cent of its funds.

The past two years, which has seen three different presidents in the institute, has also no doubt been one of some confusion and apprehension for KIST researchers.

But KIST's new president Un-Young Kim is confident he can see the right way ahead, and his views seem in tune with thinking elsewhere in the government and industry.

The government and country has tended to pressure KIST to produce visible results in a short period of time. But there seems a growing realization that KIST should turn its attention to more basic research with ap-

plications at least ten years down the road. With this in mind, the institute has just launched a new project. KIST 2000, which with an initial investment of 7.5 billion won (about \$10 million) per year from the government, aims at developing new materials, artificial limbs and organs for medical use, intelligent robots, and new informa-

undergo a similar transformation.

KIST held a symposium last month on KIST 2000 to tap outside opinions of industry and academics. The message was clear, says Kim. "We should concentrate on basic research to develop new technology and should not try and produce products."

An infusion of new blood will be essential if KIST is to succeed in its new found role. Kim says he has recruited a bright new medical researcher from Minnesota University in the United States and he plans to recruit many more from overseas, provided the government provides the necessary funds. His aim is to change the present ratio of engineers to scientists in KIST from 9 to 1 to 1 to 1 in ten years through recruitment and a process of natural attrition.

In about five years he would also like to introduce external review by foreign scientists like that carried out at KAIST. But he feels it would

be premature to do so while some KIST research still tends to consist of "imitation" of work in advanced nations. His ultimate aim is to make KIST into an institution like the Institute of Physical and Chemical Research (RIKEN) in Japan, a very worthy goal for an institute which like RIKEN is a semigovernment organization with external as well as government funding.



tion technology by the end of the decade.

The project is reminiscent of early plans for Japan's Human Frontier Science Program, with which KIST officials like to equate it. Frontiers was initially billed as a project to develop intelligent robots but subsequently was redirected into much more basic research on the brain and molecular biology. Observers hope that KIST 2000 will

derived from industrial contracts, says Kun-Mo Chung, president of the Institute for Advanced Engineering (IAE), an institute of advanced learning and research recently established by the giant Daewoo group of companies (see page 384). But the assassination of President Park in 1979 and the advent of a new administration led to a major shift in direction for KIST and other government institutes in the 1980s.

The second major step in institute building came in 1971 with the establishment of the Korea Advanced Institute of Science (KAIS), now called the Korea Advanced Institute of Science and Technology (KAIST). That sprang from a proposal in 1969 by Chung, who had spent several years in the United States at Michigan State University, Princeton University, the Massachusetts Institute of Technology, the Oak Ridge National Laboratory, the Brookhaven National Laboratory and Stanford University.

In that period, South Korea was suffering from a severe shortage of researchers with PhD and MSc degrees; students of engineering and science had virtually no choice but to go overseas to the United States, Japan, or Europe for graduate-level education. KAIS was

How does Korean science rate?

In terms of number of publications in the journals listed in the *Science Citation Index* (SCI) of the US Institute for Scientific Information (ISI), Korea comes a poor 32nd in the world with about 1,800 publications in 1990, compared with nearly 250,000 for the United States and close to 50,000 for Japan, according to statistics compiled by the Ministry of Science and Technology.

But when individual institutions are examined, the picture is brighter. KAIST produced nearly a third (28.3%) of the South Korean publications listed in SCI in 1988. And POSTECH, which was only established in 1987, published 318 in international journals in 1991. In terms of publications per faculty member, KAIST and POSTECH with about 2 papers per year in international journals are not far behind the average of 2.8 in the United States.

In terms of quality, South Korea compares poorly with seven other rapidly industrializing countries in the region (Taiwan, Hong Kong, Singapore, Malaysia, Indonesia, Thailand and the Philippines), coming next to last in terms of citation impact (average citations per paper relative to world average), according to the March 1993 issue of ISI's *Science Watch*. By field, South Korea was only strong in engineering, ranking second.

Odd as it may seem, the Philippines, which has the weakest economy and weakest industries in the region, came out top overall and top in engineering, biology and agriculture, and second in medicine. This suggests that there is not a strong link between industrial and economic strength and performance in science.

created by MOST to plug this gap.

Despite resistance from the Ministry of Education and Seoul National University, South Korea's leading national university, on the grounds that graduate level education should instead be strengthened at Seoul National University, the conservative ministry was overridden. Again the United States played a key role in advising on the establishment of KAIS. And the new institute, with substantial financial backing from MOST, offered salaries

four times those at other South Korean universities to attract top-quality faculty members from overseas, particularly Korean scientists and engineers based in the United States.

The policy succeeded. The institute is now one of the top universities in South Korea and has churned out nearly 10,000 high-quality PhDs and MScs in its short 20-year history (see below).

The 1970s were also a period of creation of heavy and chemical industries. For example, the Pohang Iron

KAIST wins praise in external review

THE Korea Advanced Institute of Science and Technology (KAIST), one of South Korea's leading universities, got glowing praise, as well as some constructive criticisms, in a review last year by the New York-based Accreditation Board for Engineering and Technology (ABET), an organization that evaluates US research institutions.

KAIST, established by the Ministry of Science and Technology (MOST) as a graduate university in 1971 and recently merged with the undergraduate school of the Korea Institute of Technology (KIT), is the only university in Korea that does not come under the stingy bureaucratic control of the Ministry of Education. (Although POSTECH, a university supported by a steel company, has almost as much freedom and comparable financial support — see next page).

From its foundation, KAIST has been well funded. In 1992, for example, the university got a total of 22 billion won (about \$30 million) in research contracts, or on average just under \$100,000 for each of the 327 faculty members. As a result, the university has been able to recruit top-quality Korean scientists and engineers from the United States. The list of universities where the present faculty members obtained their postgraduate degrees reads like a roster of the best US research institutions — MIT, Caltech, Princeton, Stanford, and Harvard to name but a few.

It is thus perhaps not surprising that

KAIST on its brand-new campus in Taeduck science town should be confident enough to invite in a team of US scientists and engineers from universities and private companies to carry out a critical review.

The ABET team led by Leslie Benmark, global systems manager of DuPont, concludes in a 76-page report after a three-day visit last September that KAIST is "a first quality institute" with "high calibre" programmes and students who are "well qualified, aggressive . . . and highly motivated". "With an increase in government support, KAIST has the potential for becoming one of the top institutions in the world".

But the undergraduate research laboratories, which originate from KIT, came in for some criticism. "A great deal of the laboratory equipment . . . is outdated or in a poor state of repair".

The reviewers also remarked on the poor English skills of the students and encouraged the university to improve English education. That should not be too difficult to do considering that many of the faculty members are fluent in English. But at present they give all their courses in Korean. KAIST could perhaps learn from some Japanese national research institutes where attempts are being made to hold

regular seminars in English because of a recognition that English is the international language of science.

KAIST, thus, has much to be proud of. But care must be taken in future recruitment of faculty. There are voices in the university saying that more faculty members should be recruited internally from the high calibre



KAIST: time to look outwards?

students in KAIST rather than recruiting from overseas. That would be a mistake.

One of the biggest weaknesses of the best Japanese national universities is in-breeding due to internal recruitment within university walls. KAIST would be well advised to maintain an open recruitment policy and to make it even more open by recruiting some non-Korean staff to maintain and strengthen its vitality.

From steel to academic excellence

POHANG was just a small fishing port and military base, in the late 1960s. Now its a bustling city boasting the world's number two steel company, and, linked to it, one of the best universities and research institutes in South Korea.

The steep descent by plane into Pohang's military airfield can be a bit hair-raising, and planes often have to divert elsewhere when the weather is bad. Outside the airport one is greeted with the traffic jams ubiquitous to South Korea as workers from Pohang Iron and Steel Company (POSCO) shuttle back and forth to work in their Hyundai cars and buses.

Hyung Sup Choi, first president of the Korea Institute of Science and Technology (KIST), says he was drafted in (somewhat reluctantly) for six months to help set up POSCO in the late 1960s when the government of then president Park decided the country needed a strong steel industry. "My only qualification was my doctoral research on the solid/air interface at the University of Minnesota", explains Choi. "So I read up the literature and we imported all the most modern steel-making technology, including continuous casting. We got raw materials from Australia and arranged the cheapest transport. Then we got lucky. Just after we opened, came the oil crisis in 1973. Steel prices leapt, and POSCO has been in profit ever since."

The question then arose as to what to do with all the profit. President Park was keen

to invest in education and research and so the idea of creating a research institute and university was born, says Choi. Pohang Institute of Science and Technology (POSTECH) and the neighbouring Research Institute of Industrial Science and Technology (RIST) opened on a plush campus in 1987.

The founders of POSTECH set out to establish a research university with low teaching loads, unlike any other university in South Korea. Over 90 per cent of the present 260 faculty, including the university's president Hogil Kim, are Koreans recruited from the best universities and institutes in the United States where many of them had spent several decades. And the university dumb-founded sceptics when on opening day it managed to attract the best students from across South Korea and shot straight to the top ranks of the nation's universities.

Free dormitory accommodation and scholarships for all of the 1,100 undergraduates and 700 graduates, plus tennis courts and a gymnasium, obviously help bring in the students. But the key attraction is a very strong research environment.

On top of a massive investment in con-

struction of the university by the steel company, POSTECH gets just over half of its annual research and development funds of 19 billion won (\$25 million) from POSCO via RIST in the form of grants for project research of interest to the company that is carried out in RIST laboratories next to the university.

But POSTECH also gets significant funding from the government and other com-



RIST's extensive laboratory complex, with POSTECH alongside.

panies as well as from the university's own foundation set up with the support of POSCO that allows research unrelated to the steel company to be carried out. For example, POSTECH has four out of the thirty KOSEF centres of excellence in university departments throughout South Korea, which are chosen by open competition and peer review and get funds of about \$1 million a year from the Ministry of Science and Technology (see table on page 383).

Much of the research in the university is clearly of direct interest to industry, for example, development of an unmanned car for Hyundai and work on robotics. But there is also substantial basic scientific research going on. One example is the study of magnetic fluxes in high-temperature superconductors in a very-well equipped low-temperature physics laboratory. There is also a very ambitious project under way to build South Korea's first synchrotron (see left).

President Hogil Kim complains that the university is hampered by bureaucratic restriction of university finances by the Ministry of Education. "For example, we can only take a 20 per cent shareholding in private companies and we can't receive tax free donations from individuals, only companies". There are also severe restrictions on the number of technical support staff, and no provisions for university overhead in the grants provided by the ministry.

This in some ways parallels the problems in universities of neighbouring Japan, which also suffer from an overbearing ministry of education. But, unlike Japan, POSTECH with the support of POSCO has achieved a considerable degree of financial autonomy from the government and Japanese universities could do well to learn from the model. □

The next leap? – a third-generation

It seems the same pioneering spirit and ambition that created the world's number two steel company on a sandy wilderness next to Pohang is driving scientists at POSTECH to build South Korea's first synchrotron. Not content to try out their hand on building a small machine, POSTECH is building a 2-GeV third-generation synchrotron using all the latest synchrotron technology, including some that has never been tested before.

Again POSCO, which is providing \$106 million of the \$181 million cost of the machine, has given the crucial backing to get the project off the ground. The Korean Science and Engineering Foundation (KOSEF) is providing the rest.

Tyong-Nyong Lee, director of the facility, apologizes for the dust when he shows visitors around. "It should not really be here but we have to be on schedule for completion at the end of next year." The 2-GeV linac leading to the synchrotron, which is "80 per cent complete" has some of the world's highest energy (80-megawatt) klystrons built by Toshiba of Japan to get electrons up to sufficient energy in the linac's short 150-metre length. There is also a considerable amount of equipment made in China as well as a 200-MW modulator built in South Korea.

Magnets for the main storage ring are just being put in place, a very delicate operation requiring precise surveying. Lee is confident everything will be complete by the end of next year and it will open to general users in 1995. But others at POSTECH worry about all the things that could go wrong with such a delicate and complex machine.

Another concern is who will use it. There are very few people in South Korea with experience in using synchrotrons and, as yet, there is very little interest from Korea industry. But Lee points out that Japan's synchrotron in Tsukuba science city initially suffered from a dearth of users but is now oversubscribed. He is hoping to attract users from overseas and he says a delegation from Australia's University of Newcastle which visited the facility last month showed great interest.



POSTECH's short but energetic linac tunnel.

and Steel Company (POSCO), now one of the world's biggest steel producers, was built from scratch at the port of Pohang by importing the latest steel-making technology from Japan and elsewhere. It has since given birth to the Pohang Institute of Science and Technology (POSTECH), one of the leading universities in South Korea, and the neighbouring Research Institute of Industrial Science and Technology (RIST), which have a unique funding relationship with the company (see page 380).

Another seed sown in the 1970s was the plan to build Taeduck science town in the outskirts of Taejeon city in the centre of the country modelled after the research triangle of North Carolina in the United States. It has taken longer than expected, but Taeduck now has about 40 government and private research institutes with nearly 12,000 scientists and engineers, comparable with the number of researchers in Japan's Tsukuba science city. It will host the Taejeon Expo (see page 377), the site of which will be absorbed into the town after the event.

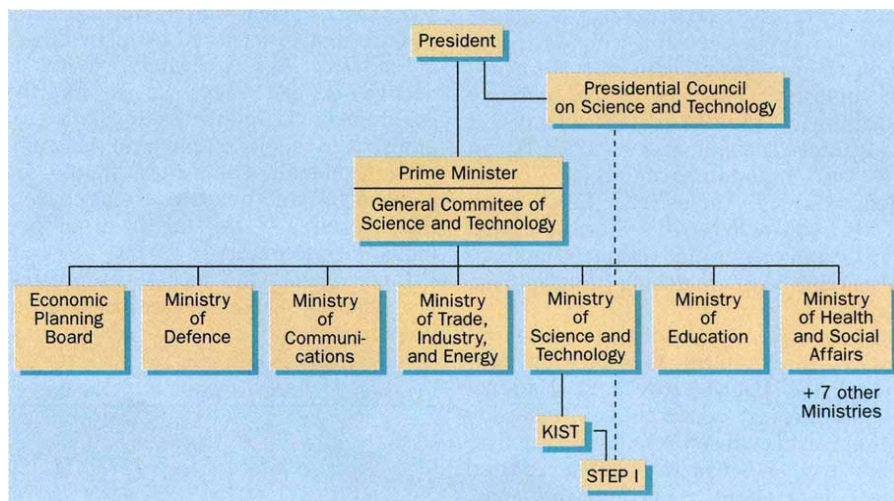
The 1980s saw an unhappy period of centralization, where government economists in the name of efficiency tried to reign in and centralize control of the newly-spawned government institutes. In 1980, KIST was merged with KAIS to create KAIST.

But the marriage of the neighbouring institutes in Seoul was always awkward and unhappy; they separated nine years later. The graduate school part of the institute (KAIS) linked up with the undergraduate school of the Korea Institute of Technology (KIT), established in 1984, to form a reborn KAIST with both graduate and undergraduate schools on a new campus in Taeduck science town.

During this period, MOST pumped money into the KIST part of KAIST; KIST researchers, lost touch with their industrial roots. Meanwhile, industry began to surge ahead in establishing its own infrastructure of research institutes. A new law for industrial technology development introduced in 1986 by the Ministry of Trade and Industry (now the Ministry of Trade, Industry and Energy) was particularly influential; company institutes began to sprout up in the late 1980s, increasing from 260 institutes with 12,500 researchers in 1986 to 1,200 and over 30,000 researchers in 1991.

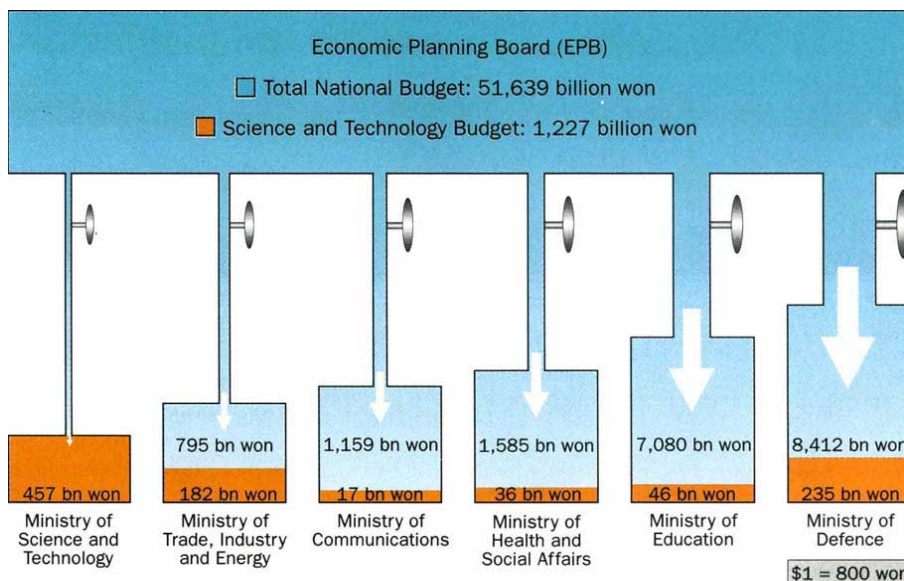
With the boom in private-sector research and growing awareness that industrial competitiveness would depend on strong industrial R&D, the powerful trade and industry ministry (and others) began to muscle in on what had

How policy is made and managed



The Presidential Council on Science and Technology is the supreme advisory body to the president. It has 11 members, including representatives of universities, government research institutes, and private industry. The present chairman is Shanghi Rhee, former minister of science and technology, and the minister of science and technology (currently Sijoong Kim) is an ex-officio member. The council meets every other week and its chairman reports to the president on a monthly basis. The General Committee of Science and Technology, which coordinates activities of the various ministries, is chaired by the prime minister. The vice chairman is the minister of the Economic Planning Board. A total of 14 ministries are represented on the committee. The Science and Technology Policy Institute (STEP), affiliated to KIST under the Ministry of Science and Technology, has recently been strengthened and in the Korean language is now called the Science and Technology Policy (and Management) Institute. STEP is responsible for managing the HAN (G-7) inter-ministry project and other national projects and has a direct reporting channel to the Ministry of Science and Technology.

How the money flowed in 1992



All budget requests are submitted to the EPB by each ministry in June/July and after screening and cutting are sent to the National Assembly for approval. But this approval is usually just a formality at least in the case of science and technology budgets. So the economists in the EPB are the key people controlling the flow of funds to the various ministries. The Ministry of Science and Technology (MOST) nominally has the largest science and technology budget. But the numbers are deceptive. For example, the budget for the Electronics and Telecommunications Research Institute (ETRI) in Taeduck science town comes under MOST but this large institute is in fact now under the control and jurisdiction of the Ministry of Communications. Also the science and technology budgets of the communications, trade, and defence ministries have been growing much faster than that of MOST in recent years.

previously been the preserve of the little science and technology ministry. KIST and university researchers came under criticism by government economists and industrialists for failing to meet the needs of industry, and the trade ministry established the Korea Academy of Industrial Technology (KAITECH) in 1989 to respond more directly to the demands of industry. This year, KAITECH will provide 90 billion won (\$115 million) for government-industry research, with matching funds from industry.

Meanwhile, responding to criticism, MOST initiated the G-7 project, now called the Highly Advanced National (HAN) Project, which aims at organizing government-industry research to catch up with the G-7 nations (Japan, United States, United Kingdom, France, Germany, Italy and Canada) in certain fields of technology, such as HDTV and broad-band integrated service and data networks, by 2001 (see *Nature* 354, 176; 1991). Again the trade ministry and other ministries have muscled in on the project and a large portion of G-7 funds are being fed to industry through KAITECH, as well as to the institutes of MOST (see table).

But industry is at best lukewarm about the project. The head of one major industrial research institute says his company already has contingency plans for the day the project fails. It's very difficult to expect anything from G-7 he says. He thinks the government should

concentrate on more basic research and leave industrial research up to companies, which have much greater resources, a view shared by many others in industry and elsewhere. What is needed instead they say is for the government to reduce its regulation of industry to allow research to blossom. "For example, companies should be allowed to issue bonds overseas to help their finances", says the head of one industrial laboratory.

Involvement in the G-7 project, nevertheless, seems pervasive both in government and industrial laboratories,

BUDGET TO CATCH UP WITH THE G-7 NATIONS

	billion won	
	1992	1993
MOST	31 (7)	50 (22)
MOTIE	11 (3)	37
MOC	10 (7)	36 (10)
ME	1	14
Sub-total	53 (17)	137 (32)
Industry	109 (27)	252 (n.a.)
Grand Total	162 (45)	389 (n.a.)

\$1 is about 800 won; n.a. not available. Figures in parentheses refer to low interest (about 6 %) loans. MOST: Ministry of Science and Technology; MOTIE: Ministry of Trade, Industry and Energy; MOC: Ministry of Communications; ME: Ministry of Environment. By 2001, government and industry plans to invest about 3,400 billion won (\$4.2 billion) and 60,000 man-years in the project.

and people do see some benefits in it, particularly in the half of the project that is devoted to long-term "fundamental technology development". SsangYong

Cement Industrial Company, for example, is carrying out research on rare-earth magnets and ferrite for use in hard-disk drives at its Taeduck research centre. "It's a way of spreading the risk and utilizing the total research resources of the whole country", says Jae-Kyun Yang, managing director and head of the company's advanced materials division.

In KIST where there was some apprehension about G-7, researchers are realizing the project can be turned to benefit. For example, Young S. Yoo of KIST's Doping Control Center is receiving G-7 support for very basic research in the hot new area of protein structure analysis. "G-7 is not entirely wrong" says the president of one university. "I think MOST dressed it up to get money from the EPB".

Chung of IAE, who served as minister of science and technology in the late 1980s, thinks that government scientists and engineers have been made "scapegoats" by government economists. He says government researchers had in fact been very loyal to the wishes of the economists by accepting the centralization and re-organization of the 1980s. "I think the days for institute building by the government are over", says Chung. Now industry has far greater manpower and money for research than the government and he says the private sector should now take the lead. It is in this vein that the Daewoo group of companies established IAE which Chung heads.

Private sector institutes flourish

Samsung Advanced Institute of Technology (SAIT) south of Seoul was one of the first and one of the biggest of many private sector institutes to open in the late 1980s. It acts as the central laboratory of the 26 companies of the Samsung group each of which have their own research institutes, some next to SAIT. Nearly half (450) of Samsung's 1,000 researchers (not including 13,000 development engineers) are at SAIT along with 150 administrators and technical support staff. The institute's president has ambitious plans to increase the total staff to 1,000 by 1995.

About 15 per cent (60) of the researchers are Koreans recruited from the United States and Japan, and Y.I. Kim, executive director for planning and administration, says that everyday he interviews two or three people who have studied in the United States or Japan and he goes on recruitment trips overseas every three months. POSTECH (see page 380) and KAIST are key sources for local recruitment.

The research and development budget for the institute is also rapidly increasing from 27 billion won last year to 40 billion won (\$50 million) this year, or about \$100,000 per researcher. The institute's goal is to carry out basic long-term research in optoelectronics, new materials, digital signal processing, mechatronics, software development, imaging and transmission technology, and development of 'environment friendly' technology, such as fuel cells, solar cells and electric car batteries. But up until now, most work has been on shorter term technology development, particularly in new materials and optoelectronics.

The Goldstar group of companies, a rival of Samsung, has also set up a central laboratory in the southern outskirts of Seoul. But the institute's vice president, Chang-Soo Kim, who used to work for DEC Corporation in the United States, does not have such ambitious goals to expand. He reckons an increase in staff from 200 to 300 in the next five years is enough. "More than 300 is too big to manage".



Samsung's fast-expanding central labs.

The Future

The area of South Korea's government research system that has been neglected throughout the past thirty years of rapid development is the university sector (with the exception of the new universities of KAIST and POSTECH). That responsibility lies with the Ministry of Education which, like its counterpart in Japan, has spent most of its energy and money on primary and high school education and very little on graduate level research.

Yet South Korea has more than 100 universities, most of them private, but the majority are very poorly equipped for research. Teaching loads for undergraduate courses are too heavy and classes too large, leaving faculty members little time for research. And the small amount of money allocated by the education ministry for research is spread too thin in an over-democratic fashion also reminiscent of Japan.

Hogil Kim, president of POSTECH, explains another fundamental deficiency of many universities this way: In 1945, there were only 11 people in Korea with first degrees in physics, but between 1945 and 1955 ten universities were established

Model for G-7 project?

with physics departments and there were no people properly qualified to fill the posts of professor. The poor quality of many faculty members in the thousands of science and engineering departments in South Korea's more than 100 universities remains a problem to this day.

Yet another problem is stifling bureaucratic control of the universities by the ministry of education. Many of the private universities are run like profit-making companies and there have been scandals involving corruption of university officials. As a result, the ministry has laid down rigid rules on the finance of universities that are applied across the board regardless of the quality of the university.

In the late 1980s, the Korea Science and Engineering Foundation (KOSEF) under MOST created an imaginative programme designed to breath new life into the best parts of the university system. KOSEF has chosen 30 centres of excellence in science and engineering departments which are provided with about \$1 million a year over a nine year period, subject to passing a review process every three years (see below). The centres are required to have research links with at least four other universities, and the engineering centres must raise several times the amount of KOSEF funds from industry.

ONE successful government-industry collaboration often quoted by supporters of the G-7 project is that to develop digital switching systems for telecommunications.

The project was launched in 1978 by the Electronics and Telecommunications Research Institute (ETRI), at that time called the Korea Telecommunications Research Institute, in collaboration with four companies – Samsung, Goldstar, Daewoo, and Oriental Telecommunications Company. The first switching systems were made by 1985 and two or three new generations of systems have been developed since with a total investment of 106 billion won (\$132 million) since 1982. The systems are not only being used in South Korea but have

Industry is also beginning to realize the need to support the universities so that they can supply industry's growing need for top-quality researchers. Companies are funding construction of new buildings in universities such as Seoul National University. And at the same university, the Research Institute for Basic Sciences of the College of Natural Sciences has recently brought together a consortium of 15 companies to support the university's research. The companies donate a total of 150 million won (\$190,000) a year that is used for symposia and to fund research of interest to the companies. Other

also been exported to Russia, the Philippines and Vietnam.

But the industrial participants no longer see much of a future for such projects. There has been constant bickering between the four partners over the sharing of the spoils. ETRI holds the patent rights, and revenues from sales are split equally between the four companies. But some of the companies feel that their contribution was larger.

Now the individual companies can mobilize far more money and manpower than ETRI. For example, ETRI's total annual budget is about \$100 million but Samsung Electronics spends \$700 million a year on telecommunications alone.

universities are likely to follow the model, says In Kyu Lee, director of the institute.

The situation in South Korea is thus better than Japan, where the ministry of education has obstructed the development of new funding sources outside the ministry. In Japan, it would be unheard of, for example, for another ministry to set up a university, as MOST did in the case of KAIST.

There is also much greater awareness in South Korea of the need for financial and administrative autonomy in research organizations. The first president of KIST Hyung-Sup Choi fought with the politicians in the National Assembly to get autonomy for KIST. "Clerks should not control research institutes" he says. He succeeded in his fight, but fortunes were reversed in the 1980s when the bureaucrats reigned in control.

The president of POSTECH has similar ambitions for autonomy. And even in industry, Chang-Soo Kim, vice-president of Goldstar's central research laboratory says his "dream" is to make the laboratory "a separate legal entity financed through royalties".

In Japan, not only are university and government researchers unaware of the need for autonomy, many don't even realize they don't have it. And when offered more financial autonomy by the government, as happened a few weeks ago in the case of several national laboratories in Tsukuba science city, the directors of the laboratories shied away, preferring to defer responsibility to the nameless bureaucracy where no individual is held accountable.

The research community in South Korea, thus, has many latent strengths over its big neighbour which need to be set loose and supported. The government needs to focus as much (if not more) effort on the improvement of the university research system as is now being devoted to short-term development of technology. The KOSEF programme

KOSEF's 30 university centres of excellence

Field	Name	University
Science Research Center (SRC)		
Mathematics	Topology & Geometry Research Center	Kyung Pook Nat. Univ.
Physics	Global Analysis Research Center	Seoul Nat. Univ.
	Center for Theoretical Physics	Seoul Nat. Univ.
	Semiconductor Physics Research Center	Chun Pook Nat. Univ.
	Research Center for Dielectric and Advanced Matter Physics	Pusan Nat. Univ.
Chemistry	Organic Chemistry Research Center	Sogang Univ.
Biology/Genetics	Center for Biofunctional Molecules	POSTECH
	Molecular Science Center	KAIST
	Research Center for Molecular Microbiology	Seoul Nat. Univ.
	Plant Molecular Biology	Kyeong Sang Nat. Univ.
	& Biotechnology Research Center	
	Research Center for Cell Differentiation	Seoul Nat. Univ.
Medicine	Research Center for New Bio-Materials in Agriculture	Seoul Nat. Univ.
	Cancer Research Center	Seoul Nat. Univ.
	Center for Mineral Resources Research	Korea Univ.
Engineering Research Center (ERC)		
Mechanics	Center for Advanced Aerospace Materials	POSTECH
Electrical engineering computer science	Center for Turbo-Power Machinery	Seoul Nat. Univ.
	Artificial Intelligence Center	KAIST
	Sensor Technology Research Center (STRC)	Kyung Pook Nat. Univ.
	Satellite Center	KIT
Material resources	Engineering Research Center for Advanced Control and Instrumentation	Seoul Nat. Univ.
	Research Center for Thin Film Fabrication and Crystal Growing of Advanced Materials	Seoul Nat. Univ.
	Rapidly Solidified Materials Research Center	Chung Nam Nat. Univ.
Chemical engineering	Material Surface Engineering Center	KAIST
	Biology Process Center	KAIST
	Research Center for Catalytic Technology	POSTECH
	Automation Research Center	POSTECH
Agriculture/fishery	Animal Resources Research Center	Kun Kuk Univ.
Medicine/pharmacology	Research Center for Ocean Industrial Development (RCOID)	National Fisheries University of Pusan
	Research Center for New Drug Development	Seoul Nat. Univ.
Nuclear engineering	New Atomic Reactor Center	KAIST

IAE — the way of the future?

The newly-opened Institute of Advanced Engineering (IAE) in Seoul is an interesting odd fish. Supported by the giant Daewoo group of companies and linked to a private university, also financed by Daewoo, IAE aims at being a world-class centre of both engineering research and graduate-level education networked with top institutions around the world.

IAE's president, Kun-Mo Chung, has had an illustrious career. Having studied for several years in leading US universities and national laboratories, he helped establish KAIST, one of Korea's leading universities (see page 379), and later served as Minister of Science and Technology. In his spare time he is South Korea's ambassador for nuclear energy cooperation in the International Atomic Energy Agency.

IAE began its first doctoral course in systems engineering in March at Ajou University south of Seoul. It also is now holding a summer school for university faculty members and company researchers at its temporary offices in Daewoo's corporate

headquarters. The first course is being given by visiting faculty from Massachusetts Institute of Technology (MIT). For example, Richard Lyon has a course on machinery noise and vibration.

Among its 14 faculty members, IAE has two foreigners (a rare breed even in South Korea's best universities), George Hazelrigg, a systems engineer from the National Science Foundation in the United States and Daniel Meneley, a nuclear engineer from Canada. IAE has just signed an agreement with the Princeton Institute of Advanced Studies in the United States to exchange personnel and to do joint research. And Chung is negotiating a similar agreement with Cambridge University in the United Kingdom.

Daewoo is pumping 500 billion won (\$600 million) into IAE over 7 years, or about the same amount of money as the total annual budget of the Ministry of Science and Technology, a sign that government really cannot compete with industry in leading industrial research these days.

needs to be expanded and better publicised. It is not well-known even in South Korea. Some research leaders in major companies have heard of the scheme, but are not quite sure where the centres of excellence are, while others have not even heard of the scheme.

Other programmes like it are also needed, but are likely to be brought into being. The Ministry of Defence, which commands a large budget for military-related research, is said to be planning to introduce a similar scheme to fund university research through its Agency for Defense Development (ADD) in Taeduck. The ministry has become more interested in civilian research since the Gulf War in which high technology played such an important role.

Also worthy of more attention is the need to internationalize South Korea's research system. The country has benefited enormously from the training of its scientists and engineers overseas, particularly in the United States, and to a lesser extent, in Japan and Europe. Most of the leaders of South Korea's best universities and research institutes have spent years out of the country. They, more than anyone, should appreciate the importance of exposure to a cosmopolitan research environment. And yet there is a hardly a non-Korean scientist or engineer to be seen in the country except those temporarily hired on high salary to be tapped for information and expertise.

As South Korea develops its own domestic sources of graduate-level scientists and as fewer students are compelled to go overseas, there is a

danger that research organizations will cease to recruit overseas and that inbreeding will set in. South Korea should not make the mistake that Japan made of temporarily opening up to foreigners and foreign ideas in the Meiji restoration last century and then turning inward for a hundred years.

Some people realize the dangers and see a new road forward. Chung's institute, which is establishing strong links with top institutes overseas, aims at being thoroughly international and has several visiting teaching staff, both temporary and permanent. Chung sees the way of the future as lying in research networks linking the best national institutes, universities and industries around the world. And, on a more local scale, this is the kind of thinking that lies behind the ambitious KOSEF centre of

excellence programme.

Much will depend on the impending restructuring of the government's management of science and technology. Discussion of restructuring is proceeding underwater after a blaze of publicity, says Man-Gi Paik, director of the industrial technology division of the Ministry of Trade, Industry and Energy. The trade ministry and other powerful ministries want MOST to revert to being an organization for coordinating rather than directing the country's research, which they say is what MOST was originally supposed to do. And one suggestion is to set up a big council of science and technology instead of MOST. Others want to strengthen the small advisory council to the president (see figure on page 381). And yet another idea is to create a new section for science and technology within the all-powerful Economic Planning Board. The Science and Technology Policy Institute affiliated to KIST, which is managing the HAN (G-7) project, may play a key role in the restructured system.

The Ministry of Education, which has recently been embroiled in corruption scandals exposed by the new administration has been weakened by the resignation of several top bureaucrats and is unlikely to have a strong say in the reorganization being put together by a restructuring committee and economic advisors to the president.

The outcome of the restructuring should become clear in September in the next session of the National Assembly when a new law for government re-organization will be debated. Whatever route is chosen, a strong organization independent of the battling ministries and the institutes and universities they oversee seems necessary in this new age where so many government ministries have become deeply involved in science and technology policy. □

ACKNOWLEDGEMENTS

The author, David Swinbanks, Publisher and Tokyo Correspondent for *Nature* in Japan, would like to thank Young Sik Kim, director of administration at KIST, and Chongsae Lee advisor to KIST, for their invaluable help in setting up interviews and arranging transport to collect material for this special issue during the period 23 June to 3 July. Thanks are also due to their colleagues Hyung-Jin Jung, vice president of KIST and the members of KIST's office of public relations and international cooperation.

The special issue was inspired by a suggestion of Kim Jin Hyun, then minister of science and technology, during an earlier visit to South Korea by the author in May 1992. And the author would also like to thank the many people not listed here who contributed during four previous visits to South Korea between November 1991 and October 1992.

In addition to those people mentioned in the text, the following people (listed in order met) gave generously of their time during the present visit:

Y.M. Jung, President, Samsung Advanced Institute of Technology (SAIT)
Seong Soo Jeong, Team Manager of Public Affairs, SAIT
Shanghi Rhee, Chairman, Presidential Council on Science and Technology
Hoagy Kim, Secretary General, President Council on Science and Technology

Young-Il Youm, Dean of Academic Affairs, POSTECH
Dong Han Kim, Head of Dept. of Chemistry, POSTECH
Sang-Il Choi, Professor of Physics, POSTECH
Sung-Ik Lee, Assoc. Professor of Physics, POSTECH
Hyun Jae Jo, Public Information Dept., POSTECH
Chan Juh Rhee, Vice President, RIST
Dal-Woo Kim, Head of Physics Group, RIST
Sung-Chul Shin, Director Public and International Relations, and Associate Professor, Dept. of Physics, KAIST
Zeungnam Bien, Chancellor and Dean of Academic Affairs, KAIST

Kim Yong Chae, Public Information Office ETRI
In Seok Han, Supervisor of Protocol and Secretarial Group, ETRI

Usung Chung, Director, EXPO '93 Press Centre
Si Joong Kim, Minister of Science and Technology
Youn Soo Sohn, Assistant Minister, R & D Policy and Coordination Office, MOST

Moon Hi Han, Division Director, STEPI
Chulsu Kim, Minister of Trade, Industry and Energy
Chemin Rim, Chief Secretary to Minister, Ministry of Trade, Industry and Energy

Dong-Baek Lee, Executive Vice President, SsangYong Cement Industrial Co Ltd.

In Key Moon, General Manager PR Team, SsangYong Cement Industrial Co., Ltd.

Endless ripples on the sands of time

The formation of ripple patterns on sandy beaches is more than an inconvenience for those who walk on them barefoot, but a challenge in physical analysis that should occupy people for seasons to come.

As much of the population of the Northern Hemisphere heads for the discomforts of beach recreation, with sand grains in every sandwich and every sunburn scar, the lesser discomfort of walking barefoot over apparently flat sand that is cryptically marked out with ripples at intervals of 10 cm or so may have been overlooked. It is therefore relevant, and certainly a public service, that two Japanese physicists have now produced an explanation of why the apparently flat surface of a beach is often not flat at all, but instead marked with an array of hard and potentially punishing ripples. The calculation also has the interest of being a hybrid between a lattice calculation and one appropriate to a system with continuous variables, which means that it may even be read when the beach season is over.

Evidently the place to start is with an understanding of the physics of the transport of sand grains along a beach. The foundation work in the field is a monograph by R. A. Bagnold published in 1941 under the title *The Physics of Blown Sand and Desert Dunes*. If memory is an accurate guide, the book has several then-remarkable photographs of dunes in the Sahara.

There are two ways in which a sand grain can move. It can roll along the surface under the force of the wind, which Bagnold called creep. And it can move by saltation, or jumping, which requires that it should first be rolling and should then collide with another sand grain in such a way that it is carried into the air, accelerated by the wind and finally deposited on the surface with momentum that may be enough to set other sand grains rolling. An extension of the second mechanism, saltation, is what Bagnold called suspension; the wind is strong enough, or the lift of the sand grain otherwise sufficient, for it to be carried almost indefinitely far.

The new analysis of ripple patterns on the beach is by Hiraku Nishimori and Noriyuki Ouchi of Ibaraki University, near the Pacific coast a few hundred kilometres north of Tokyo. Their first concern is to model the process of saltation (*Phys. Rev. Lett.* **71**, 197; 1993). The process is conceptually simple. Suppose the surface of the sand is described by a pair of Cartesian coordinates x and y , that the wind is blowing in the x direction and that the height of the surface at every point in the plane is some value of $h(x,y)$, which is naturally a function of position in the plane. Each

single saltation movement of a sand grain will therefore reduce the height of the surface by some quantity q at the point from which the grain departs and increase it by the same height at some distance L downwind.

So far, so good. The trouble is that L conceals a welter of physics. For one thing, the greater the height from which a sand grain is carried away, the greater will be the distance it travels, or the saltation length. The characteristics of the wind near the surface of the beach will also matter, but the authors insist that the controlling parameter is not the velocity of the wind, but the shear-stress of the wind, linked with the vertical gradient of the wind velocity.

For those with time to reflect (perhaps in between sand-contaminated sandwiches) there are even further complications. The saltation length will also obviously also depend on the position on the profile of a ripple pattern from which a sand grain is carried. To put it in an extreme form, a sand grain lifted from the top of a ripple will travel further than one lifted off from the windward face (which is likely to stick a little further up the same slope). Those who enjoy constructing interlocking sets of non-linear differential equations will relish the elaborations that can be constructed; solving them, of course, is another matter.

The next step is to allow for what happens when a sand grain lands after a saltation step. In reality, it is more than likely that it, or one of its neighbours, will continue moving in the direction of the wind. For the sake of simplicity, the authors neglect this process (which Bagnold reckoned to be unimportant), but they do allow for the way in which a sand grain settling on a surface will bed down under gravity, pushing aside to some extent its nearest and next-nearest neighbours.

How to solve this set of equations? Only by computer, one would guess. So the authors dutifully describe a series of simulations of the behaviour of figurative sand grains on a lattice, in each case starting with a random profile. By supposing that $L = L_0 + bh(x,y)$, where L_0 and b are constants (the first a function of wind characteristics), they conclude that ripple patterns do indeed form naturally. More than that, there is a linear relationship between the parameter L_0 and the wavelength of the ripples: the greater the

wind force, the greater the wavelength. But below a certain value of L_0 , no ripples form. Similarly, too fast a relaxation under gravity of the profile of the surface is the death of ripples.

But as it happens, something can be said and done analytically. Drawing on an unpublished report dating from 1951 at the University of Tokyo, the authors outline how to set up equations to describe the dynamics of the increase and decrease of the height of the sand profile at any point of the surface. It is merely necessary to allow that the rate of increase of height at any point is largely the difference between the rates at which sand is added and removed. The former is the rate at which sand is removed a saltation length away upwind, the latter the same rate at the point in question.

The equations cannot be solved explicitly, but it is possible to use them to tell whether particular ripple patterns will be stable with the passage of time. Among other things, this line of argument leads to the conclusion that there must, indeed, be a wind speed below which stable ripple patterns do not exist.

By the same test, the arc-shaped profiles of desert dunes emerge from the simulations (but, reading between the lines, only fitfully), and are also suggested by the stability analysis. Here, of course, it is necessary explicitly to allow for the way in which a sand grain carried off from the windward face of a dune is more than likely to embed itself higher on the same face (which may account for the nearly vertical profile near the top of the windward face of a sand dune).

It is unlikely that even Nishimori and Ouchi will wish to claim that their investigation is complete. On the contrary, it has all the virtues of an open-ended research programme. There is plenty of scope for adding to these simple models hydrodynamic calculations of the behaviour of wind near the surfaces of rippled sand, of a first principles calculation of the degree to which wind will lift a sand grain and, while Bagnold may have declared rolling creep to be less important than saltation, that was half a century ago...

In short, there is no reason why those of a sufficiently enquiring disposition should return from incarceration on the summer's beaches complaining that they were bored. There is plenty to be done, this summer and next.

John Maddox

Menopause for thought

Linda Partridge

In the twilight of their careers eminent scientists not infrequently enter a philosophy, in which academic activity shifts gradually from the laboratory bench to the armchair. The menopause of the human female is both much more abrupt and much harder to understand. Simply put, natural selection is expected to cause the evolution of life histories that maximize the number of surviving progeny left by individuals. Why then does the fertility of women cease totally at a time of life when life expectancy is still high, thereby seeming to reduce lifetime reproductive success? In a paper in *Evolutionary Ecology*¹, Alan Rogers has revisited this question, bringing an age-structured theory of kin selection to bear on it.

Cessation of fertility at the menopause is a consequence of failure of the ovaries to produce the egg-bearing follicles. Follicle number declines and they become less responsive to gonadotrophic hormones. Hormonal changes such as the drop in oestrogen level are a secondary consequence of this primary reproductive failure. The unique feature of the human menopause is the specific shut-down of fertility at a time when other physiological systems continue to function.

One explanation of the menopause is that it is an artefact of the high survival rates in modern human societies; it would not have been seen in the harsher circumstances under which the human life history evolved, because death would have already intervened. However, demographic data do not support this idea². Furthermore, it explains neither why a menopause is absent in other mammals, including chimps and gorillas, when survival rates are increased in benign conditions, nor why it is absent in human males.

The idea that the menopause could have evolved by natural selection was explored informally by G. C. Williams³. In pre-agricultural societies, death in childbirth would have been common, and its probability would have increased with the age of the mother. Human juveniles are also dependent on their mother for their survival for several years after birth, and her ability to care for them would decline with her age. Under these circumstances, it might pay a woman of 45 or 50 to devote the whole of her declining energies to the care of her dependent children and grandchildren, and to avoid putting them at risk of losing her through the increasing hazards of childbirth. Her loss of fertility would therefore be compensated by the increased survival or fertility of her existing descendants, and the menopause would evolve as a natural

contraceptive. An alternative method would have been behavioural avoidance of mating but, given the predilections of males, a physiological method may have been safer.

Arguments such as these, despite their intuitive appeal, can be a misleading guide to the course of evolution. What Rogers has done is to formalize Williams's thinking, and make explicit the circumstances under which a new, rare allele causing continued fertility at the normal age of menopause would invade the population. The model allows the allele to have effects on the fertility and survival both of the mother and of her descendants, and extends existing models of kin selection to include effects at different ages and also the occurrence of time-delays between the cost of the act of altruism (maternal sterility with onset at menopause) and the benefit to the recipient (increased survival or fertility of descendants).

The first type of allele considered was one that caused continued fertility at the usual age of menopause, at the cost of an increasing risk of maternal death during childbirth. If maternal death occurred, then the survival and fertility of existing dependants was impaired. The second allele involved a different potential cost of continuing fertility, namely reduced ability to care for existing children during pregnancy and immediately after childbirth, exacerbated by the increasing age of the mother.

Using data from the 1906 population of Taiwan, an agricultural society for which good demographic data are available, Rogers evaluated the action of natural selection on alleles of these two kinds. Female fertility declined markedly with age before the menopause in this population, and it followed that most children of women who reached menopause were already of reproductive age, while grandchildren were still pre-reproductive. Care was therefore likely to benefit mainly the fertility of offspring and the survival of grandchildren.

Rogers shows that mortality in childbirth increasing with maternal age cannot account for the evolutionary stability of the menopause. Using an estimate of probability of death during childbirth of 1 in a 100, and assuming that maternal care doubled the fertility of children and resulted in 100 per cent survival of children and grandchildren until they were 10, menopause was still strongly selected against. These assumptions were all stacked towards finding a benefit for the menopause, and even a tenfold increase in maternal mortality in childbirth with increasing age left the conclusion un-

altered, so it seems robust. The peculiar hazards of human birth therefore seem unlikely to account for the existence of the menopause.

Next, Rogers considered the adverse effects of continuing fertility on care of existing dependants. Again stacking the assumptions in favour of menopause, and in particular assuming that adverse effects on maternal care of other descendants lasted until a new child was three years old, menopause became marginally evolutionarily stable. But with more realistic assumptions it seems likely that the benefit would be reversed. At first sight, therefore, the evolution of menopause by natural selection seems implausible.

An advantage of a theoretical model is that it makes assumptions explicit, and encourages measurement of real values of the important variables. What we notably lack for models of the evolution of the menopause is any direct information on what would happen to maternal survival and fertility if it did not occur. Several factors were not considered in Rogers's model, as he himself points out.

Increases in maternal mortality during birth and in the loss of care to dependent offspring because of the arrival of a new child may not be the only considerations, and a number of other processes could favour the menopause. The likelihood of death of women increases after the menopause because of ageing, in the absence of birth of new children. If women remained fertile at menopause, then each new child would anyway become progressively less likely to survive because of increasing likelihood of maternal death at times other than during childbirth. Continuing fertility might also cause a more rapid acceleration of maternal mortality rates. Furthermore, fertility has already declined at the time menopause occurs, so that its abolition is less costly than it would be at a younger age.

Finally, pregnancy and very young children are more physically demanding for mothers than are older children. It seems probable that the main adverse effect of increasing maternal age on parental care would be on fetuses and very young children rather than on older ones, again tilting the evolutionary odds towards menopause. During their evolution other mammals and human males may not have displayed the protracted, costly and widely dispensed care of descendants seen in human females, and it may be this difference that accounts for the unique occurrence of the menopause. □

Linda Partridge is in the Institute of Cell, Animal and Population Biology, University of Edinburgh, Ashworth Building, West Mains Road, Edinburgh EH9 3JT, UK.

1. Rogers, A. R. *Evol. Ecol.* **7**, 406–420 (1993).

2. Hamilton, W. D. J. *theor. Biol.* **12**, 12–45 (1966).

3. Williams, G. C. *Evolution* **11**, 398–411 (1957).

Exocytosis goes with a SNAP

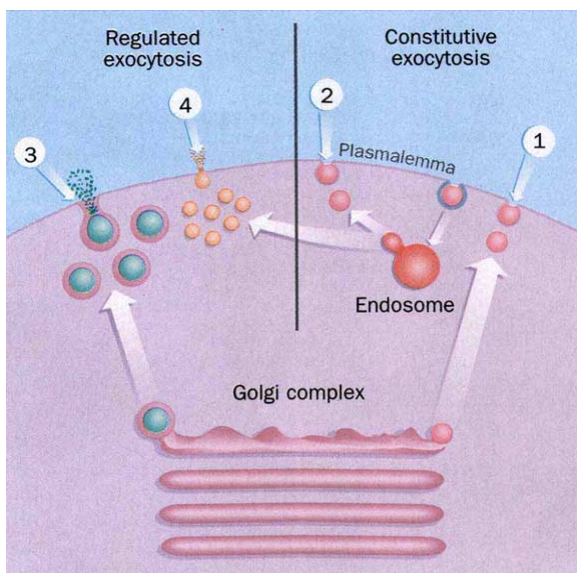
Pietro De Camilli

THE regulated fusion of synaptic vesicles with the plasmalemma is a highly specialized process. Yet it is becoming increasingly clear that the molecular mechanisms of intracellular membrane fusion are highly conserved throughout evolution and at all levels of the secretory pathway. On page 445 of this issue¹, Osen-Sand *et al.* now provide evidence to support the hypothesis that key components of the fusion machinery are involved in both regulated and constitutive exocytosis (see figure for an explanation of the two processes). They show that SNAP-25, a protein implicated in the fusion of synaptic vesicles with the plasmalemma, is also involved in constitutive axonal growth.

SNAP-25 (for synaptosomal-associated protein of M_r 25K) was first characterized in 1989 as the product encoded by a neuron-specific mouse messenger RNA². It is a membrane-associated cytoplasmic protein of 206 amino acids, and is highly conserved from *Drosophila* to mammals (D. Larhamman, personal communication). It is widely distributed in the brain — although levels of expression vary in different neuronal populations — and it is concentrated in nerve terminals where a large proportion of the total protein seems to be localized in the plasma membrane. From its sequence, SNAP-25 is a hydrophilic protein. But in the adult nervous system it is palmitylated at one or more of its four, closely spaced cysteines clustered in the region between amino acids 84 and 92 (the 'cysteine quartet'), and it behaves as an integral membrane protein²⁻⁴.

SNAP-25 has two isoforms, a and b, which are generated by alternative splicing and are differentially expressed during development. SNAP-25a is expressed early on, whereas SNAP-25b expression begins with the onset of synaptogenesis and it becomes by far the predominant isoform in the adult brain. The two isoforms are generated by the alternative use of two homologous exons which encode 39 amino acids, including the cysteine-rich motif. In SNAP-25a, the highly conserved cysteine quartet of SNAP-25b is replaced by a slightly different sequence which may affect the protein's fatty-acylation³. Because SNAP-25 immunoreactivity has a diffuse distribution in axons before synaptogenesis, it was proposed that the switch between the two

isoforms correlates with a different subcellular distribution^{3,5} (M. Wilson, personal communication). This would be consistent with a role of SNAP-25 palmitylation in its targeting to, or retention in, nerve terminals, because other peripheral membrane proteins concentrated in nerve terminals — GAP-43 (ref.



Exocytosis is the process by which an intracellular vesicle fuses with the plasmalemma. As a result, the vesicle contents are released into the extracellular medium (secretion) and the components of the vesicle membrane become part of the plasmalemma. Some vesicles fuse with the plasmalemma constitutively, whereas others accumulate under the plasmalemma and fuse only after an appropriate stimulus, often a rise in cytosolic Ca^{2+} . Constitutive exocytosis is used by cells for non-regulated protein secretion and for insertion of newly synthesized protein components into the plasmalemma (1). Endosome-derived vesicular carriers which shuttle plasmalemma components (such as transferrin receptors) back to the cell surface also undergo constitutive exocytosis (2). Regulated exocytosis is characteristic of secretory granules derived from the Golgi complex, which store concentrated secretory products (3), and of specialized endosome-derived vesicular carriers such as synaptic vesicles (4). There is increasing evidence that a similar mechanism may underlie these various types of exocytosis; as discussed in the text, proteins encoded by the SNAP-25 gene family seem to be an important part of the fusion machinery.

6) and GAD-65 (ref. 7), for example — are palmitylated on cysteines.

The concentration of SNAP-25 in nerve terminals led to the suggestion that it may participate in the docking or fusion (or both) of synaptic vesicles with the plasmalemma². Two sets of observations strongly support this hypothesis.

First, Söller *et al.* found that SNAP-25, together with the plasmalemma protein syntaxin, participates in the formation of a complex with the synaptic vesicle protein synaptobrevin II. The formation of this

complex requires ATP and cytosolic proteins — *N*-ethylmaleimide-sensitive factor (NSF) and the so-called 'soluble NSF attachment proteins' (SNAPs), an acronym coincidentally similar to SNAP-25 — which are generally concerned in intracellular fusion⁸. That this complex is essential for synaptic vesicle exocytosis is supported by the observation that tetanus and botulinus toxin B block neurotransmitter release by cleaving one of its components, synaptobrevin^{9,10}.

Second, SNAP-25 was shown to be homologous to part of the yeast *SEC9* gene product (Sec9p), a protein of 74K (651 amino acids) which is required for the fusion of Golgi-derived secretory vesicles with the yeast plasma membrane. SNAP-25 is similar to the carboxy-terminal 231 amino-acid residues of Sec9p (19 per cent identity and 60 per cent similarity), and expression of the tail region of the Sec9p from a low copy yeast plasmid is enough to prevent the lethality of $\Delta sec9$ mutants and restore wild-type secretory function. Strikingly, Sec9p does not contain any cysteine (V. Bankaitis *et al.*, personal communication). This observation counts against the hypothesis that SNAP-25 or related proteins may act as acceptors for fatty-acyl groups during the fusion reaction⁸.

These two findings not only support the idea that SNAP-25 is central to synaptic vesicle exocytosis, but also suggest that SNAP-25-related proteins are generally involved in exocytosis from all cells. Osen-Sand *et al.*¹ now provide an interesting validation of this second hypothesis. Using an antisense oligonucleotide approach, the authors looked into the function of SNAP-25 in developing neurons before synapse formation. When antisense-oligonucleotides (complementary to coding regions common to SNAP-25a and b) were added to cortical neurons in primary culture, they reduced SNAP-25 levels by as much as 94 per cent. At the same time axonal outgrowth decreased, as did expression of the

synaptic vesicle protein synaptophysin (but not of other neuronal proteins), possibly indicating a reduction in the potential to form presynaptic sites. When added to PC12 cells before treatment with NGF, the oligonucleotides decreased SNAP-25 expression by more than 75 per cent; and they blocked NGF-induced neurite elongation without affecting outgrowth of small spikes, including those induced by cyclic AMP. Finally, reduction of SNAP-25 levels in the retinae of chick embryos, induced by injection of anti-

sense oligonucleotides in the eye, resulted in impaired axonal growth from retinal neurons and in the reduction of the thickness of the inner plexiform layer.

Although synaptic vesicle exocytosis and recycling occurs throughout the distal arbor of developing axons¹¹, axonal elongation is likely to involve constitutive exocytosis at the growth cone of a distinct class of vesicles¹. So implicating SNAP-25 in axonal elongation is in agreement with its having a function independent of synaptic vesicle exocytosis. It will be of interest to determine whether SNAP-25a is involved in axonal elongation, while SNAP-25b, which is dramatically up-regulated during synapse formation, is involved in synaptic vesicle exocytosis.

Does SNAP-25 depletion affect exocytosis from neuronal cell bodies and dendrites? Osen-Sand *et al.*¹ do not report any obvious effect of this sort. Yet, in all cells of higher eukaryotes, exocytosis is required not only for growth and secretion, but also for the very active exocytocytic recycling that takes place at the cell surface. Given that SNAP-25 has a homologue in yeast, SNAP-25 homo-

logues are likely to carry out house-keeping functions in all cells; a SNAP-25 isoform related (or identical) to non-neuronal isoforms of SNAP-25 may have this function in the cell body of neurons depleted in SNAP-25.

The characterization of the SNAP-25 gene family, and of the proteins' precise mechanism of action, are the next tasks to be tackled. Given the speed of advance, we should not have to wait long for them to be accomplished. □

Pietro De Camilli is in the Department of Cell Biology and the Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, Connecticut 06510, USA.

1. Osen-Sand, A. *et al.* *Nature* **364**, 445–448 (1993).
2. Oyler, G. A. *et al.* *J. Cell Biol.* **109**, 3039–3052 (1989).
3. Bark, I. C. *J. molec. Biol.* (in the press).
4. Hess, D. T., Slater, T. M., Wilson, M. C. & Pate Skene, J. H. *J. Neurosci.* **12**, 4634–4641 (1992).
5. Oyler, G. A., Polli, J. W., Wilson, M. C. & Billingsley, M. L. *Proc. natn. Acad. Sci. U.S.A.* **88**, 5247–5251 (1991).
6. Liu, Y. *et al.* *Neuron* **6**, 411–420 (1991).
7. Christgau, S. *et al.* *J. Cell Biol.* **118**, 309 (1992).
8. Söllner, T. *et al.* *Nature* **362**, 318–323 (1993).
9. Schiavo, G. *et al.* *Nature* **359**, 832–835 (1992).
10. Link, E. *et al.* *Biochem. biophys. Res. Commun.* **189**, 1017–1023 (1992).
11. Matteoli, M., Takel, K., Perin, M. S., Südhof, T. C. & De Camilli, P. *J. Cell Biol.* **117**, 849–861 (1992).

GLOBULAR CLUSTERS

Next swing of the pendulum

Charles D. Bailyn

IN the middle of the 1970s, when X-ray sources were discovered in globular star clusters, there was considerable excitement surrounding the possibility that these sources might contain black holes. Ten years later, it was generally agreed that this interpretation was false. Now, in a new swing of the pendulum, two papers on pages 421 and 423 of this issue^{1,2} suggest that globular clusters may indeed contain black holes, albeit of much smaller mass than those envisaged in the 1970s.

The original suggestion was that the cluster X-ray sources could be explained by a black hole several hundred times the mass of the Sun, sitting in the cluster centre accreting gas and stars from its dense surroundings. Theoretical studies of cluster evolution which indicated that the core of a cluster should collapse to infinite density in a finite time lent credence to this idea. But the X-ray sources could also be explained as binary star systems in which a neutron star accreted material from its companion. Supporting this hypothesis was the clear evidence of binarity in several similar X-ray sources located outside clusters. The significantly greater frequency of X-ray sources in clusters compared with the rest of the galaxy was explained as a consequence of 'tidal capture', a process in which stars in a dense cluster transfer sufficient kinetic energy into tidal oscillations to

create tight binary systems.

Two lines of evidence showed the latter hypothesis to be preferable. First, the mysterious X-ray bursts observed in the cluster sources were convincingly modelled as thermonuclear explosions of material accreted onto the surface of a neutron star³. Accretion onto black holes would not create densities high enough to initiate the bursts, as the holes lack a surface on which material can accumulate. Secondly, X-ray observations with high spatial resolution demonstrated that the X-ray sources were not coincident with the cluster centre⁴, as would be expected of black holes significantly more massive than an individual star. Indeed, the spatial distribution of the sources was found to be compatible with binary systems of two solar masses, as predicted by the neutron star model. In the face of this contrary evidence, the initial enthusiasm for black holes diminished greatly.

Interest in stellar-sized black holes in general has been heightened recently by both observational and theoretical developments. Observationally, it has been shown that a number of transient X-ray sources outside globular clusters contain compact objects of greater than three times the mass of the Sun^{5–7}. Basic considerations of general relativity imply that these objects must be black holes. The boom in supernova modelling following

the observation of the supernova in the Large Magellanic Cloud in 1987 and the success of supernova search programmes has reinforced the idea that some supernova explosions can leave behind black holes^{8–9}. Extrapolations from these results has led to an estimate of a population of over a hundred million stellar-sized black holes in our Galaxy alone¹⁰.

Now, Kulkarni and his collaborators¹ and Sigurdsson and Hernquist² have explored the consequences of the less massive black holes in globular clusters. Both groups demonstrate that the holes, being more massive than a typical cluster member, will rapidly sink to the centre of the cluster, where they will interact with each other. These interactions are likely to result in the expulsion from the cluster of most if not all of the holes on a relatively short timescale. If a hole or two remains behind, however, there may be interesting observational consequences, particularly if, as seems likely, they form binary pairs.

Kulkarni and collaborators suggest that binaries comprising a black hole and a normal star may result in a transient X-ray-emitting system, just as they do outside clusters. Sigurdsson and Hernquist point out that binary systems containing two black holes may have a devastating effect on the surrounding normal stars, stripping material from them, and in some cases disrupting them completely. These effects may, they say, account for recently observed anomalies in the stellar populations at the centres of the densest clusters¹¹.

Thus we have come full circle since the 1970s. At that time, unexplained observations prompted the theoretical speculation that globular clusters might contain black holes. Now, theoretical results are suggesting observational tests of the presence of a rather different black hole population. Direct or indirect observational confirmation would give us a fascinating new window on the complex problems of the evolution of globular clusters, as well as on the origin of the black holes themselves. □

Charles Bailyn is in the Department of Astronomy, Yale University, New Haven, Connecticut 06511, USA.

1. Kulkarni, S. R., Hut, P. & McMillan, S. *Nature* **364**, 421–423 (1993).
2. Sigurdsson, S. & Hernquist, L. *Nature* **364**, 423–425 (1993).
3. Ayasli, A. & Joss, P. C. *Astrophys. J.* **256**, 637–665 (1982).
4. Grindlay, J. E. *et al.* *Astrophys. J.* **282**, L13–L16 (1984).
5. McClintock, J. E. & Remillard, R. A. *Astrophys. J.* **308**, 110–122 (1986).
6. Casares, J., Charles, P. A. & Naylor, T. *Nature* **355**, 614–617 (1992).
7. Remillard, R. A., McClintock, J. E. & Bailyn, C. D. *Astrophys. J.* **399**, L145–L149 (1992).
8. Wooley, S. E. & Weaver, T. A. A. *Rev. Astr. Astrophys.* **24**, 205–254 (1986).
9. Maeder, A. *Astr. Astrophys.* **264**, 105–120 (1992).
10. van den Heuvel, E. P. J. in *Proc. Satellite Symp. 3: Space Sciences with Particular Emphasis on High-Energy Astrophysics* 29–36 (ESA, 1992).
11. Djorgovski, S. *et al.* *Astrophys. J.* **372**, L41–L44 (1991).

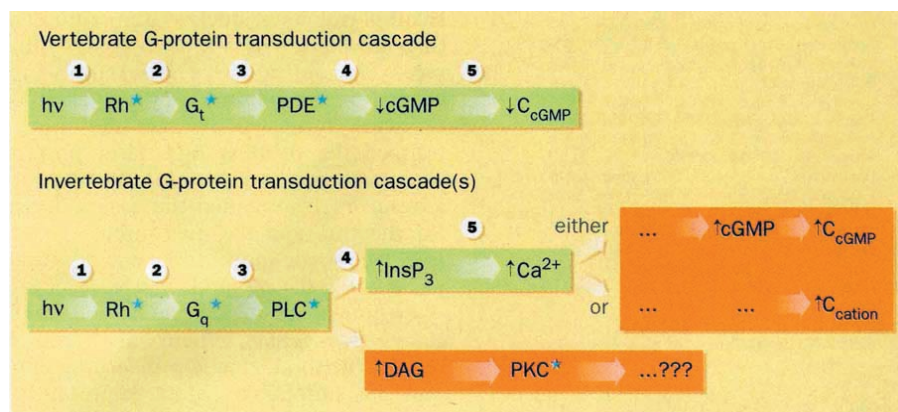
Strange case of the third eye

Edward N. Pugh, John Lisman and Richard Payne

THE first intracellular recordings from photoreceptors were made in an invertebrate, the horseshoe crab *Limulus*, in the 1950s¹. They showed that light depolarized the membrane and set expectations about how other photoreceptors would behave. So it came as a surprise when, several years later, Tomita convincingly demonstrated that the light response of fish cone photoreceptors is hyperpolarizing². The finding was confirmed in

vertebrates a cyclic GMP-phosphodiesterase, which causes cytoplasmic cGMP to decline and cGMP-gated channels to close; the consequent reduction in Na⁺ and Ca²⁺ influx hyperpolarizes the membrane.

In invertebrates, the enzyme activated by the G protein is phospholipase C (PLC); elimination of this enzyme, as occurs in the *Drosophila* mutant, *norpA*, completely abolishes the response to light⁴. One consequence of the activation



Top, the five major steps in activation in the vertebrate G-protein phototransduction cascade: 1, activation of rhodopsin (Rh) by photoisomerization; 2, catalytic activation of G-protein (G_t, transducin) by Rh*; 3, activation of phosphodiesterase (PDE) by binding of G_t; 4, hydrolysis of cGMP by PDE*; 5, closure of cGMP-gated channels. * denotes an activated protein. ↑, ↓, the direction of a concentration change or change in conductance of an ion channel. Below, steps 1 and 2 of the invertebrate cascade are similar to those in the vertebrate, with an activated rhodopsin catalysing the activation of a distinct G-protein (G_q). PLC, phospholipase C; InsP₃, inositol trisphosphate; DAG, diacylglycerol. Steps 1 to 5 are thought to occur in all invertebrate photoreceptors (green bands). The remaining steps are not well established (red bands): one possible path leads to activation of a cGMP-gated conductance; another, to the opening of a nonspecific cation conductance. The DAG branch of the PLC cascade is thought to be involved in inactivation (see text).

numerous subsequent intracellular recordings, and by the 1970s the surprise became the rule — vertebrate photoreceptors hyperpolarize.

Now come Solessio and Engbretson, who on page 442 of this issue³ report the first exception to this rule. Recording from morphologically typical vertebrate photoreceptors of the lizard parietal eye (the so-called third eye), they find a primary light response that is depolarizing. These same photoreceptors also have a hyperpolarizing light response, which is spectrally distinct from the primary response, and is exhibited only at high intensities. These remarkable findings raise many questions.

One concerns the molecular mechanism of the depolarizing response. Phototransduction in both invertebrates and vertebrates begins when light is absorbed by rhodopsin-like photopigments, which are a subclass of the superfamily of G-protein-activating receptors. In vertebrate rods and cones, the G protein acti-

of PLC is the production of inositol-trisphosphate (InsP₃), which triggers release of Ca²⁺ from intracellular stores⁵. A second consequence is the production of diacylglycerol, which activates protein kinase C. The link between Ca²⁺ release or diacylglycerol production, and the opening of depolarizing cation channels in the plasma membrane, remains unclear. Surprisingly, channels in excised patches of *Limulus* photoreceptors, which resemble the light-activated channels seen in cell-attached patches, can be activated only by cGMP⁶. So one possibility is that Ca²⁺ released by InsP₃ leads to an increase in cGMP, which opens the channels. Alternatively, the opening of cation channels may be linked to the depletion of the intracellular calcium stores whose release is triggered by InsP₃ (ref. 7).

Given the central role of PLC in invertebrate phototransduction, it will be fascinating to see if PLC is involved in the depolarizing response of the lizard pariet-

RESUME

Tread alert

THE increase in concentrations of stratospheric sulphates, a possible cause of ozone destruction, could be due to human sources of carbonyl sulphide. The gas is concentrated in urban areas, suggesting that it might originate from motor vehicles. But the amount of OCS in car exhaust is insignificant. Now W. H. Pos and H. Berresheim (*Geophys. Res. Lett.* **20**, 815–817; 1993) have discovered another automotive source which could be more important. By studying the emission of OCS from tyre treads under conditions simulating road surfaces in summer, they estimate that tyre wear could account for about 6.7 per cent of the total OCS flux into the atmosphere. Should environmentalists be lobbying against hand-brake turns as well as aerosol cans?

Air traffic control

RADAR has been used to watch the sky for migrating insects for nearly 25 years. But bulky equipment and arduous analysis have meant that it hasn't been used for the kind of long-term monitoring for which it is ideally suited, and which would yield results of economic importance, particularly in agriculture and in developing countries. That should no longer be the case, thanks to advances described by A. D. Smith *et al.* (*Phil. Trans. R. Soc. B* **340**, 393–404; 1993). The speed, direction, orientation, size and shape of a flying insect can be discerned from the way it modulates a beam of a 'vertical-looking' radar: it should be possible to work out a radar 'signature' for individual species. The data can be processed by an inexpensive personal computer, and a little modification should allow the continuous monitoring of large-scale insect migrations by a distributed computer network.

Magnetic appeal

THE thinnest magnets around have to be monolayer ferromagnetic films. M. Wuttig *et al.* may now have identified a new class of ultrathin magnetic materials — ordered surface alloys (*Phys. Rev. Lett.* **70**, 3619–3622; 1993). On depositing enough manganese to give roughly half a monolayer's coverage on a nonmagnetic copper substrate, they found an orderly pattern in which manganese replaced every other surface copper atom. Although manganese and copper are of similar atomic diameter, the manganese atoms buckled outwards from the surface by about 0.3 Å. Magnetism, say the authors, is probably what stabilizes this corrugated surface over a broad temperature range: their calculations indicate that the exchange interaction (coupling between electron spins) between electrons in the manganese *d*-orbitals gives a net energy saving if the higher orbitals are populated, producing, in effect, a fatter atom.

al eye. A demonstration that the PLC pathway plays such a role in the vertebrate parietal eye might help to elucidate its puzzling presence in retinal photoreceptors. There have been several reports of light-activated PLC in vertebrate photoreceptors⁸, and a very recent report that bovine cones express a homologue of the *norpA* gene⁹. The diacylglycerol branch of the PLC pathway activates a C kinase, one target of which is rhodopsin itself¹⁰. Interestingly, in invertebrates, C kinase seems to be involved in reactions that terminate the light response^{11,12}.

A second intriguing issue raised by the new work is the basis of the hyperpolarizing response seen at high light intensities. Is the lizard parietal-eye photopigment bistable (like that of the invertebrates) or does its photoproduct decay (like that of the other vertebrate pigments)? If the pigment is bistable, then the bright lights that produce hyperpolarization might do so by simply converting active photoproduct back to rhodopsin, as is the case in invertebrates. Such back conversion could reduce the level of activation. Alternatively, the hyperpolarization might be mediated by a distinct molecular pathway, perhaps the

conventional vertebrate cascade (see figure). However things turn out, photoreceptors of the lizard parietal eye are going to break more of the established 'rules' of vertebrate phototransduction. □

Edward N. Pugh is in the Department of Psychology and Institute of Neurological Sciences, University of Pennsylvania, 3815 Walnut Street, Philadelphia, Pennsylvania 19104, USA. John Lisman is in the Department of Biology, Brandeis University, 415 South Street, Waltham, Massachusetts 02254, USA. Richard Payne is in the Department of Zoology, University of Maryland, College Park, Maryland 20742, USA.

1. Hartline, H. K., Wagner, H. G. & MacNichol, E. F. *Cold Spring Harb. Symp. quant. Biol.* **17**, 125–141 (1952).
2. Tomita, T. *Cold Spring Harb. Symp. quant. Biol.* **30**, 559–566 (1965).
3. Solessio, E. & Engbretson, G. A. *Nature* **364**, 442 (1993).
4. Pak, W. L. *et al. Science* **194**, 956–959 (1976).
5. Payne, R., Corson, D. W., Fein, A. & Berridge, M. J. *J. gen. Physiol.* **88**, 127–142 (1986).
6. Bacigalupo, J., Johnson, E. C., Vergara, C. & Lisman, J. *Proc. natn. Acad. Sci. U.S.A.* **88**, 7938–7942 (1991).
7. Hardie, R. C. & Minke, B. *Neuron* **8**, 643–651 (1992).
8. Ghalayini, A. J. *Neurochem.* **57**, 1406–1412 (1991).
9. Ferraira, P. A., Shortridge, R. D. & Pak, W. L. *Proc. natn. Acad. Sci. U.S.A.* **90**, 6042–6046 (1993).
10. Newton, A. C. & Williams, D. S. *J. biol. Chem.* (in the press).
11. Smith, D. P. *et al. Science* **254**, 1478–1484 (1991).
12. Hardie, R. C. *et al. Nature* **363**, 634–636 (1993).

MOLECULAR MOTORS

Wrestling with kinesin

Jonathon Howard

Two groups — one reporting in *Science*¹, the other on page 457 of this issue² — provide the first measurements of the single-motor force generated by the microtubule-based motor protein kinesin. The techniques used by the two sets of authors are completely different. Unfortunately, so too are the answers!

To understand how motor proteins such as myosin, dynein and kinesin generate force, it is necessary to measure that force; after all, force is the important product of the reaction. But to measure the force, one must literally wrestle with the proteins. One hand must hold down the motor, while the other pulls on the filament against which the motor is working. Clearly, one cannot measure force when the proteins are freely diffusing in solution. Nor can one do it in a crystal where the correct orientation is lost. So most of the techniques of biochemistry simply won't do. Over the past several years the race has been on to develop methods for measuring force at the single molecule level. The new papers are the first reports of those endeavours. But the results are too different to be reconciled; more experiments are needed and the limitations of the two approaches must be scrutinized.

Motor proteins are enzymes that con-

vert the chemical energy derived from the hydrolysis of the γ -phosphate bond of ATP to mechanical work used to power cellular motility. Most of our knowledge of this reaction comes from the study of muscle contraction. The precise, almost crystalline array of filaments in an intact muscle fibre makes this a beautiful preparation for mechanical studies. But the large number of myosin molecules in a muscle fibre (about a billion) makes it difficult to extrapolate results down to the single molecule level. The development of *in vitro* motility assays, which permit the study of motility by a small number of motors, has partially filled this gap between the muscle and the molecule. In these assays, a surface is coated with the purified motor protein and the movement of the associated filaments across the surface in the presence of ATP can be imaged under the microscope. But measurement of the force generated by single myosin molecules is still tricky; apparently, tens of molecules are necessary for continuous myosin motility³.

Answering the biophysicists' need for a motor that operates solo, single kinesin molecules have been shown to be sufficient for microtubule-based motility *in vitro*^{4,5}. Kinesin was first identified as a protein that transports organelles along

microtubules within neurons, and it is the founding member of a superfamily of eukaryotic proteins whose members also power mitosis and meiosis⁶. Unlike myosin, which operates in large, ordered arrays *in vivo*, kinesin probably operates alone or in small numbers to move small objects around the cell. So it is well suited for single-molecule mechanical studies. But the experiments are still not easy as the forces are minute: if one divides the maximum force generated by skeletal muscle by the number of myosin molecules operating in parallel, one obtains a force of 2–3 piconewtons (pN) (ref. 7).

To measure force, Kuo and Sheetz¹ used optical tweezers to immobilize a small plastic bead stuck to a microtubule that moved across a kinesin-coated surface. The optical force arises from the change in momentum of light as it is refracted by the bead⁸. Kuo and Sheetz trapped the bead at high laser power. They then reduced the intensity until the kinesin force exceeded the escape force for the trap and the microtubule broke free. The single-motor force was reported to be 1.9 ± 0.4 pN (s.d., $n=18$).

Although the magnitude of the force seems reasonable, there are several reasons to think that the measurements may be unreliable. Most importantly, optical tweezers, being such new instruments, are difficult to calibrate. The force measurements were made when the optically trapped bead was very close (much less than a wavelength) to a glass surface whose refractive index differs from that of the solution. But the calibrations were done some distance from the surface. The electromagnetic fields are different in the two cases and we expect a difference in the trapping forces. A second problem is that the calibrations were made on unconstrained beads, whereas in the measurements the beads were constrained by the glass surface and perhaps by the microtubule as well. These constraints are likely to displace the bead vertically from the centre of the trap, so reducing the horizontal trapping force⁸. The bottom line is that the calibration must be done under the exact experimental conditions.

Other questions remain. Why does the motor force change threefold over the course of the measurement? Is this consistent with the small coefficient of variation reported for the entire population of motors? And why are Kuo and Sheetz's results qualitatively different from those of Block *et al.*⁵, who found that the optical force promoted detachment of the motors rather than the cessation of movement?

The centrifugal microscope⁹ used by Hall *et al.*² provides a completely different means of measuring force. These workers challenged kinesin to move a demembrated sperm cell in a microscope whose stage could be rotated at speeds up to 5,000 r.p.m. The size and density of the

sperm head are sufficiently great that inertial forces in the piconewton range can be achieved. The results are very surprising: at low kinesin density on the surface, where it is assumed (but not shown) that the motion is due to single molecules, the inertial force required to stop the movement completely is only 0.12 ± 0.03 pN.

This force seems too small; indeed, it is too small even to overcome the viscous drag force experienced in the absence of a centrifugal force. Perhaps even more worrisome is that the force at high kinesin density is only 0.4 pN, even though we might expect that tens or hundreds of motors are at work. This high-density force is perhaps three orders of magnitude smaller than the force inferred from the bending of microtubules in *in vitro* motility assays^{10–12}. A possible explanation for the low apparent force is that something other than the inertial force alone is stopping the motors. Consistent with this idea, rotation of the microscope stage slows and stops sperm moving in all directions, even sperm moving in the same direction as the centrifugal force.

The wrestling match between motor proteins and biologists is set to go many rounds yet. Several other groups are developing alternative methods for measuring forces. Furthermore, force is itself only one half of the equation. The hope is to measure simultaneously the displacement of individual motors as they hydrolyse single molecules of ATP and work against known loads. In this way the energetics of the reaction can be understood and compared to the biochemistry.

Like the engines that humans build, motor proteins must be tailored to the particular job that they perform. The mechanical behaviour of cells is perhaps as complex and coordinated as their electrical or chemical activity — given the diversity of these mechanical behaviours, we can expect to discover a rich set of design principles used in the construction of molecular machines¹³. □

Jonathon Howard is in the Department of Physiology and Biophysics, SJ-40, University of Washington, Seattle, Washington 98195, USA.

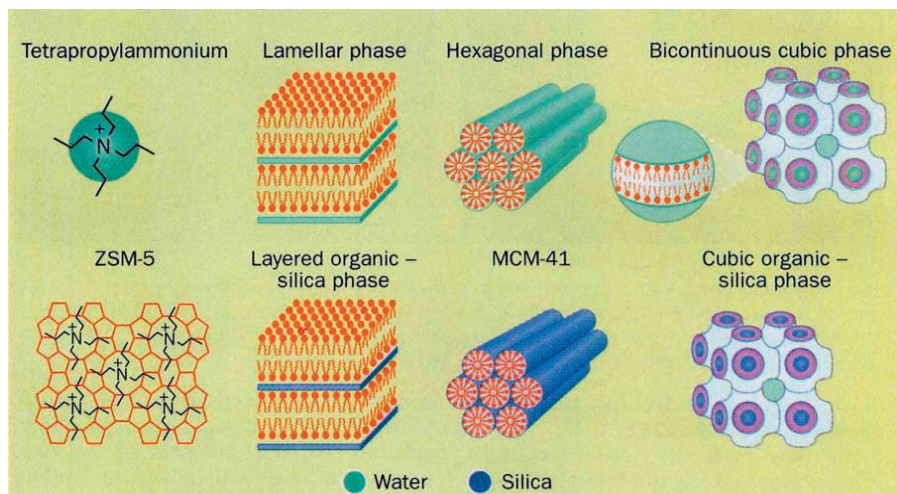
Organizing for better synthesis

Mark E. Davis

To sieve out one molecular species from another, reach for a zeolite; or now that they are available, choose a microporous inorganic oxide tailor-made for the job by taking a miniature 'plaster cast' of assemblies of organic molecules. On page 430 of this issue, Archibald and Mann¹ show how this method can be stretched to produce microtubules or stacks of lamellar crystals.

ognition (see figure of zeolite ZSM-5⁹). Nonetheless, selected crystalline structures can be made this way with the ability to recognize or organize molecules of 0.5–1.5 nm in size.

In the new method, the inorganic oxides are organized by self-assembled organic aggregates held together only weakly. Here the organization is at a much larger length scale than obtainable with single



Just some of the geometries to be obtained by 'templating'. Top, the molecule or molecular aggregates; bottom, the zeolite or microporous oxide produced.

Zeolite and zeolite-like molecular sieves, with macroscopic properties closely connected to their atomic structure, have attracted particular attention in fields where molecular recognition is needed, such as catalysis, separations and chemical sensing^{2,3}. Chemists wishing to extend the range of geometries available have attempted to do so by growing zeolites on 'template' molecules, but success has been achieved only on small-scale structures, with pores up to 1.2 nm across. The use of intricate patterns of self-assembled organic molecular aggregates^{4–8}, as exemplified by the results of Archibald and Mann, has now led to the development of inorganic structures with organization on the length scale of 2–100 nm.

The new synthetic methods differ considerably from the earlier 'templating' method. With zeolites, single organic molecules are used to arrange the inorganic oxide species into units with a suitable geometry to combine and make the synthetic material⁹ (see, for example, the accompanying article on the zeolite DAF-1). It is templating only in a loose sense, as there is no exact correspondence between the form of the template molecule and the final product, certainly none comparable to that found in biological molecular recognition.

The sort of self-assembled structures formed in aqueous solutions of amphiphiles are illustrated in the figure.

Each of these molecular aggregates contains an aqueous region that is in intimate contact with the organic phase. For example, the hexagonal phase consists of close-packed cylinders where the hydrophobic alkyl chains are gathered in the centre of the cylinders and the polar groups are in contact with the continuous aqueous region that lies between the cylinders. Addition of silicate species, such as silicon alkoxides, into a synthesis mixture containing aqueous amphiphiles yields silicon oxide species throughout the aqueous region. In some syntheses, the amphiphiles apparently self-assemble and then organize the silicate species within the aqueous regions, whereas in others, it is clear that the presence of the silicate species act in concert with the amphiphiles during the self-assembly process^{6,10}. Whatever the mechanism, the results are fascinating: composite structures (organic-silica) have been synthesized with lamellar^{4,7}, hexagonal (MCM-41)^{5,6,10} and cubic^{5,6} topologies (see figure).

Once the mechanisms that govern the self-assembly of these organic-inorganic composites have been worked out, the

1. Kuo, S. C. & Sheetz, M. P. *Science* **260**, 232–234 (1993).
2. Hall, K., Cole, D. G., Yeh, Y., Scholey, J. M. & Baskin, R. J. *Nature* **364**, 457–459 (1993).
3. Uyeda, T. Q. P. *et al.* *Nature* **352**, 307–311 (1991).
4. Howard, J., Hudspeth, A. J. & Vale, R. D. *Nature* **342**, 154–158 (1989).
5. Block, S. M., Goldstein, L. S. B. & Schnapp, B. J. *Nature* **348**, 348–352 (1990).
6. Endow, S. A. & Titus, M. A. *Rev. Cell. Biol.* **8**, 29–66 (1992).
7. Lombardi, V., Piazzesi, G. & Linari, M. *Nature* **355**, 638–641 (1992).
8. Ashkin, A. *Biophys. J.* **61**, 569–582 (1992).
9. Ojima, K., Chaen, S., Kamitsubo, E., Shimmen, T. & Sugi, H. *Proc. natn. Acad. Sci. U.S.A.* **87**, 7893–7897 (1990).
10. Brokaw, C. J. *Cell Motil. Cytoskeleton* **6**, 347 (1986).
11. Amos, L. A. & Amos, W. B. *J. Cell Sci. Suppl.* **14**, 95–101 (1991).
12. Gittes, F. *et al.* *J. Cell Biol.* **120**, 923–934 (1993).
13. Leibler, S. & Huse, D. A. *J. Cell Biol.* **121**, 1357 (1993).

A double-barrelled zeolite

TEMPLATING on lipid aggregates and their ilk may be the new trend, as Davis's News and Views article describes, but there is life in the old molecular templating method yet — the remarkable zeolite created by J. M. Thomas and colleagues at the Royal Institution testifies to that

pores, 1.6 nm across and large enough to accommodate multiply branched alkanes or other large molecules. The A channels are linked to each other through shorter 10-ring pores, and to the B channels by 8-ring pores (free diameter ~ 0.39 nm). The organic 'template' used in this

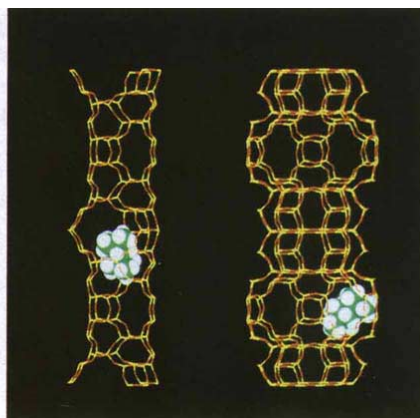
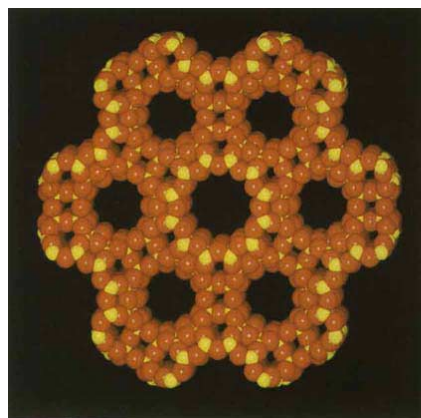
in the channels. The template can be removed by calcination in dry oxygen at 610 °C to yield the acidic form of the material, thus opening up the possibility that it may be used in acid-catalysed reactions.

The structure looks much like a hybrid between the two most widely used zeolitic catalysts, zeolite Y (with the structure of the natural mineral faujasite) and zeolite ZSM-5, shown in Davis's figure. Zeolite Y, used for catalytic cracking in gasoline manufacture, has a three-dimensional network of pores in which 12-ring windows (free aperture ~ 0.74 nm) open into larger cavities, ~ 1.2 nm across. ZSM-5, which is used in xylene isomerization and methanol conversion, contains two different channels of similar size to the A channels in DAF-1, but in this case the channels intersect. Because of its family resemblance to zeolites Y and ZSM-5, it is expected that DAF-1 should also be useful for shape-selective hydrocarbon conversions.

But that is not all. The complex architecture of non-intersecting pores offers, at last, a real prospect of molecular 'traffic control' (E. G. Deruano, *J. Catal.* 100, 541–544; 1986). The hope is that two types of large molecules, such as the branched alkanes essential to the petrochemical industry, could be preferentially accommodated in the A or B channels. Small molecules, such as methanol, would be free to nip between the channels and react with the imprisoned molecular juggernauts.

Anthony K. Cheetham

Anthony K. Cheetham is in the Materials and Chemistry Departments at the University of California, Santa Barbara, California 93106, USA.



The hexagonal crystal structure of DAF-1, showing a space-filling model viewed down the unique axis with the B channel (at the centre) surrounded by a hexagon of A channels, and side views of the A (left) and B (right) channels, with iso-octane molecules located at their respective energy minima (courtesy of P. A. Wright).

(P. A. Wright *et al.* *J. chem. Soc. chem. Commun.* 633–635; 1993).

The new material, known as DAF-1 (Davy–Faraday-1), is the first zeolite with two large, but different, non-intersecting channels (see figure). Each of the channels has 'windows' made up of 12 tetrahedrally coordinated aluminium or phosphorus atoms, linked by oxygen bridges. The smaller channel (A) has a uniform free diameter of only ~ 0.61 nm because of puckering of these 12-rings, whilst the wider channel (B), of free diameter 0.74 nm, opens into bulbous

case is decamethonium hydroxide, $[(\text{CH}_2)_5\text{NMe}_3]_2(\text{OH})_2$, which is incorporated into the channels of the final product. In chemical terms, the system is a metal-containing aluminium phosphate, otherwise known as a MeAPO molecular sieve. The metal (magnesium, for example) substitutes for aluminium in the aluminium phosphate framework (E. M. Flanigen *et al.* in *New Developments in Zeolite Science and Technology*, 103–112 (Kodansha, Tokyo, 1986)) so charge-compensating cations, in this case the template, are required

possibility of designing new materials will really be with us. Chemistry libraries overflow with information on synthesizing specific organic microstructures, whether by employing new amphiphiles, counterions, cosurfactants or mixtures thereof. With a little ingenuity, we should be able to adapt this to the design of composite (organic–inorganic) structures.

In fact, even lipid tubules can be used to organize silicon⁸ and iron¹ oxides. As Archibald and Mann show, lipid-iron oxide composites can be synthesized as tubes or lamellar disks depending on the synthesis conditions¹, showing how flexible this new method can be. Moreover, the inorganic portion need not be inorganic oxides. Clever choices of inorganic precursors should allow a broad spectrum of inorganic materials to be formed.

The concept of self-assembled microstructures serving as structure-directing agents, new as it is, is not limited to inorganic materials. For example, biconti-

nous phases can be used to organize organic species which are then photopolymerized; after extraction of unpolymerized components, what are left are porous organic monoliths^{11–13}. These materials, with pores in the size range 2–50 nm, are useful for separations technology, and for soft contact lenses and other biological and medical applications. These materials can readily be made biocompatible, as they are in many ways similar to biological membranes. In fact, biomineralization in living organisms^{14,15} may well resemble this organization of inorganic materials by self-assembled molecular aggregates.

The importance of these new synthetic efforts is that three-dimensional structures with prescribed architectures — regularly spaced pores, thin sheets or scaffolding-style structures — containing order on the 2–100 nm length scale can be formed through self-assembly processes. Order on this size offers opportunities for

synthesizing novel catalysts, electronic materials and the like. For example, the material MCM-41^{5,6,10} would be formed from the hexagonal phase illustrated in the figure where the aqueous region is

1. Archibald, D. D. & Mann, S. *Nature* **364**, 430–433 (1993).
2. Davis, M. E. *Ind. Engng Chem. Res.* **30**, 1675–1683 (1991).
3. Ozin, G. A., Kuperman, A. & Stein, A. *Angew. Chem.* **101**, 373–390 (1989).
4. Sakata, K. & Kunitake, T. *J. chem. Soc. chem. Commun.* 504–505 (1990).
5. Kresge, C. T. *et al.* *Nature* **359**, 710–712 (1992).
6. Beck, J. S. *et al.* *J. Am. chem. Soc.* **114**, 10834–10843 (1992).
7. Dubois, M., Gulik-Krzywicki, Th. & Cabane, B. *Langmuir* **9**, 673–680 (1993).
8. Baral, S. & Schoen, P. *Chem. Mater.* **5**, 145–147 (1993).
9. Davis, M. E. & Lobo, R. F. *Chem. Mater.* **4**, 756–768 (1992).
10. Chen, C. Y. *et al.* *Microporous Materials* (in the press).
11. Anderson, D. M. *World Patent Appl.* WO 90/07575 (1990).
12. Anderson, D. M. & Strom, P. *Physica A* **176**, 151–167 (1991).
13. Drummond, C. J. & Ninham, B. W. *Chem. Austr.* 529–531 (1992).
14. Heuer, A. H. *et al.* *Science* **255**, 1098–1105 (1992).
15. Mann, S. & Perry, C. C. *Adv. inorg. Chem.* **36**, 137–200 (1991).

replaced with silica. On removal of the amphiphiles, uniformly sized pores of 2 to 10 nm (the size is fixed at one value in this range depending on the method of preparation) are left. This pore size range is appropriate for separating biomolecules or constructing quantum-confined materials within the pores for electro-optical applications.

Finally, a distinction should be made between zeolites (and zeolite-like molecular sieves) and the new materials described above. It lies in the local atomic structures of the inorganic regions. The inorganic oxide structure of a zeolite has a repeat unit and a unit cell. The self-assembled organic-inorganic composites do not contain crystalline inorganic re-

gions. Rather, their inorganic portions have local order (first and maybe second coordination) similar to that of amorphous oxides¹⁰. So their physicochemical properties are most likely to resemble those of amorphous oxides rather than zeolites. For example, aluminosilicate MCM-41 has acid properties similar to amorphous aluminosilicates¹⁰. Will self-assembled, organic, molecular aggregates be found that can organize inorganic species into three-dimensional zeolite-like crystals? □

Mark E. Davis is in the Chemical Engineering Department, California Institute of Technology, Pasadena, California 91125, USA.

TECTONICS

Margins declassified

John C. Mutter

SOME years ago I eagerly participated in an activity that is common in the Earth sciences, but that I have grown to regret. It involved a sort of taxonomy of passive continental margins (margins at which there is no subduction)¹. The idea was that two distinct classes can be recognized; one is said to be 'volcanic', the other 'non-volcanic'. Within each class there were certainly differences; in fact, every margin examined seemed to hold some special characteristic that made it almost unique, while still falling comfortably within family lines. On page 433 of this issue, Holbrook and Keleman² give us cause to ask why we engaged in this classification exercise, and whether it really led to a better understanding of the processes that shape continental margins.

At the time it seemed there were good reasons to establish a classification or naming scheme for passive margins. These margins result from the rending of continental lithosphere, and the creation of new oceanic lithosphere in the intervening region. Seismic reflection imaging to mid-crustal depths on many passive margins showed large, rotated crustal blocks, bound by normal faults, the geometry of which could be used to estimate the amount of upper crustal stretching³.

Exceptions

But there were aberrant examples that did not show the classically defined rotated fault block structures. Instead their structure often included a single, large wedge-shaped body tilted seaward, within which individual reflectors were upwardly arched, nearly horizontal in their upper levels and becoming steeper down-dip. One of the earliest recognized dipping reflector sequences occurs on the conjugate continental margins of the

Norwegian-Greenland sea⁴. Scientific drilling legs on the Norwegian margin established that the uppermost part of the dipping wedge consisted of basaltic lavas, and that by inference the entire wedges were built of an extrusive complex⁵. These seaward-dipping reflector sequences were very easy to recognize in seismic reflection images, especially as their host margins seemed not to subside as deeply as others, and hence tended to accumulate less sediment (which muffles the reflections).

The final piece of evidence that seemed to establish the volcanic nature of such margins came from seismic refraction measurements. These gave estimates of the velocity structure of the deep crust in regions that are typically difficult to image by reflection. Beneath the dipping wedges, a distinct layer was often found with a seismic velocity of around 7.2 km s⁻¹ that was not represented in either the crust of the adjacent continent or the crust of the ocean basin⁶. These high-velocity layers were interpreted to be formed of igneous rocks that intruded and/or underplated the thinned continental crust of the margin. That is, they were the intrusive and plutonic counterpart of the extrusive sequences that formed the dipping reflectors.

The story was complete, and a simple classification followed: margins with dipping reflectors were named volcanic margins; those with rotated fault blocks and no dipping reflectors were named non-volcanic margins. What we then required were two explanations.

Rifting of continental lithosphere can be thought of as akin to sea-floor spreading in that the lithosphere separates and underlying mantle rises upward to fill the void. On rising, the mantle decompresses and hence melts. At mid-ocean ridges,

decompression melting is responsible for essentially all the magma produced to form oceanic crust. At very slow spreading rates the upwelling velocity is necessarily slow, and although decompression melting must still occur, the volume of melt produced is smaller as there is time for conductive cooling. This is what is believed to happen under 'normal' circumstances in passive margin evolution; some rift-related melting probably takes place but generally not enough to leave a recognizable signature.

Hotspots

One way to create additional magma during rifting is to add heat to the system so that decompression melting begins at greater depths. White and McKenzie⁷ estimated how much magma would be generated for given amounts of extension and showed that the observed thicknesses of igneous rocks could be produced — if the mantle could be made hot enough. Small perturbations on normal mantle temperatures proved inadequate, so they suggested hotspots (plumes of hot material from the deep mantle) as the heat source, citing the influence of the Icelandic hotspot on rifting in the Norwegian-Greenland sea.

They further proposed that hotspots influenced rifting at many continental margins, pointing to examples of hotspot trails leading away from margins where significant outpourings of lava occurred at the time of rifting, producing 'flood basalt' provinces such as the Parana and Etendeka, the Deccan and Rajmahal traps. Most of the hotspot-influenced margins were also known to include seaward-dipping reflector sequences, and some even had the high-velocity lower crustal layers. A process was attached to a descriptor and 'volcanic' margins became synonymous with hotspot margins.

This appealingly simple conjecture always had one or two problems. One was that the distribution of 'volcanic' sequences indicated by dipping events did not seem to know where the hotspot was located. They did not get progressively thinner with distance from the hotspot, being almost uniformly developed along large segments of individual margins. Furthermore the effect of the inferred hotspots seemed to turn off very quickly after breakup; crust of normal thickness floors ocean basins immediately next to volcanic margins. Another problem was timing — why should hotspots come to life exactly at the time of rifting? A few places were also known where dipping reflector sequences were present, but no hotspot accompanied rifting. These awkward features were difficult to encompass within the hotspot model, but did not prevent its widespread adoption.

Holbrook and Keleman may now have dug up a really thorny problem. The

passive margin of the United States East Coast has long been regarded as the head of the household in the 'non-volcanic' family. The series of seismic studies that Holbrook and Keleman summarize shows that it actually has the appropriate attributes — dipping reflectors, high-velocity deep layer — for a volcanic margin.

Why has it taken so long to recognize these features? Well, only the most recent studies have used seismic sources powerful enough to penetrate the huge absorbing barrier of sediment on the margin. Also, because the margin lacks any evidence of flood basalt sequences or a hotspot trail associated with breakup, no one knew where to look for dipping reflectors. Hinz⁸ had actually suggested their presence in older, poorer reflection data, so perhaps we just didn't want to believe that the sequences were there because they did not fit our preconceptions. Now we are forced to conclude that the United States East Coast cannot belong to the 'volcanic' family if the hotspot mechanism is to be invoked, nor yet to the 'non-volcanic' family because it lacks the proper family resemblance.

Things may be even worse. Holbrook and Keleman estimate that the average seismic velocity in the lower crust could be as high as 7.5 km s^{-1} . A velocity of 7.2 km s^{-1} is possible in magma derived from partial melts⁷ because higher mantle temperatures produce more melting, and this forms magma with more of the dense magnesium oxide component. An increase of 300°C — a reasonable anomaly for a hotspot — gives a seismic velocity of 7.2 km s^{-1} . But extrapolating these results, we find that a leap of more than 550°C is required to reach 7.5 km s^{-1} . Such a temperature perturbation would produce almost twice the thickness of igneous crust estimated to be present on the East Coast margin. If it resulted from a hotspot, surely so fierce a hotspot would leave a huge trail?

Perhaps we can account for the high-velocity lower crust once we know just what rocks are present at depth in the crust; ultramafic rocks (those rich in iron and magnesium) can exhibit near-mantle velocities. More troublesome for models of margin evolution is the apparent lack of continental crust in the structure sections for the United States East Coast². Dipping reflector sequences lie in contact with the intrusive/plutonic section, with nothing in between. Either the continental crust is so thin or so intruded that it cannot be recognized, or it simply isn't there. The transition from clearly continental crust to a region with no continental character is completed in $\sim 20 \text{ km}$, half the crust's original thickness. Lithospheric rupture has apparently erased any vestige of continental crust to produce one of the most extraordinary geological boundaries on Earth. This boundary is, to my mind, as

critical to understanding the process of lithospheric rupture as the volcanic sequences themselves.

The action of a hotspot cannot reasonably be blamed for this boundary. It is as if the continental lithosphere that was broken here was extraordinarily weak. Anderson and others⁹ have suggested that continental flood basalts occur when rifting of very thick lithosphere causes mantle upwelling from unusual depths. If thick lithosphere includes a thick crust it may indeed be much weaker than thin lithosphere and rupture without a protracted period of extension or the formation of a wide, distended margin. For rapid separation, decompression melting would not be suppressed and, in fact, might be increased if convection due to lateral thermal gradients enhances the mantle flow. A regionally warm mantle, rather than an isolated hotspot, would further weaken the lithosphere throughout the extensional province, depress the solidus and reduce mantle viscosity, to give rapid upwelling along the entire margin. Holbrook and Keleman point out that this would lead to melting at initially low melt fractions and high pressure, as has been seen in basalts from the North Atlantic Tertiary province¹⁰.

Declassification

So where do we stand? It seems to me that the whole idea of making a strict bipolar classification of margins was an error. By labelling margins 'volcanic' and 'non-volcanic', we thought we had brought an understanding to them that we had not.

Surely the Earth does not behave this way. Surely tectonic stresses, combined with the strength of the lithosphere and the condition of the underlying mantle, will give rise to an essentially continuous range of responses and a corresponding variety of margin forms. Our challenge is not to classify the observable properties of the system, but to derive from them a closed set of parameters and physical laws that properly describe its behaviour. □

John C. Mutter is in the Department of Geological Sciences, Columbia University, and Lamont-Doherty Earth Observatory, Palisades, New York 10964, USA.

1. Mutter, J. C., Buck, W. R. & Zehnder, C. M. *J. geophys. Res.* **93**, 1031–1048 (1988).
2. Holbrook, W. S. & Keleman, P. B. *Nature* **364**, 433–436 (1993).
3. Beaumont, C., Keen, C. E. & Boutlier, R. *Geophys. J. R. astr. Soc.* **70**, 667–715 (1982).
4. Mutter, J. C., Talwani, M. & Stoffa, P. L. *Geology* **10**, 353–357 (1982).
5. Eldholm, O. *et al. Nature* **319**, 360–361 (1986).
6. LASE Study Group *Mar. Petrol. Geol.* **3**, 234–242 (1986).
7. White, R. S. & McKenzie, D. *J. geophys. Res.* **94**, 7685–7729 (1989).
8. Hinz, K. *Geol. Jb.* **22**, 5–28 (1981).
9. Anderson, D. L., Tanimoto, T. & Zhang, Y. *Science* **256**, 1645–1651 (1992).
10. Fram, M. S. & Leshner, C. E. *Nature* **363**, 712–715 (1993).

Silver linings

Of the sunlight that hits the Earth, about 30 per cent is just reflected away again, largely from white cloud. The rest warms the Earth, and potentially contributes to the greenhouse effect. One way of countering this menace would be to increase the Earth's reflectivity. This journal has mooted some ways of doing this, such as covering the Sahara desert with mirrors or putting vast sunshades into orbit, which would embarrass even Daedalus. He has a far neater solution: increase the whiteness of the clouds.

Thanks to their surface tension, small cloud droplets have a higher vapour pressure than big ones. So small droplets tend to evaporate, while big ones take up the released vapour and grow even bigger. This slow replacement of many small droplets by a few bigger ones reduces the reflectivity of the cloud. Daedalus reckons that by adding detergent to a young cloud, you could reduce the surface tension of its droplets. They would cease to evaporate or merge into bigger ones, so the cloud would stay highly reflective. His new bubble-blowing machine releases thousands of buoyant, hydrogen-filled detergent bubbles every minute. Launched into young, growing clouds, the benevolent bubbles will dope them with detergent and stabilize them at their brightest and best.

A bubble barrage will not merely keep a cloud white. It will keep it in being. Normally, most clouds don't last long. Their droplets steadily merge and agglomerate and precipitate as rain. But detergent in a droplet can do more than merely reduce its vapour pressure. By covering the droplet with a monolayer of non-volatile surface-active molecules, it can cut its rate of evaporation almost to zero. Even better, two such coated droplets will not merge if they happen to collide; they will simply bounce apart again. (You can see this effect in the skating droplets often thrown up in detergent-laden washing-up water.) A detergent-doped cloud could neither evaporate nor rain. It should drift around the atmosphere almost permanently, reflecting sunlight away into space.

Thus the greenhouse effect will be brought under fine control. Daedalus's bubble barrages will load the atmosphere with an increasing population of highly reflective, permanent clouds. By reflecting so much light away on top, permanent clouds will seem dark and rain-charged from underneath. But with luck they will look odd in some way, so that we will soon come to recognize them as rainless frauds. Weather everywhere will get cloudier and colder, but not rainier.

David Jones

Gold in South Wales coal

SIR — We have found geochemically significant concentrations of particulate gold in coals from the South Wales coalfield. We think this is the first described record of native gold in coal which demonstrates direct precipitation of the gold at low temperatures. The few previously reported modern instances of anomalously high gold values have been clearly shown to be of detrital origin, related to a neighbouring gold source^{1,2}.

The gold was initially found as small (2–10 µm) collomorphic grains associated with carbonates (Ca, Mg, Fe, Mn), sulphates (Ca and Ba), clays (kaolinite and illites), sulphides (Fe, Co, Ni, Cu, Zn, Mo, Cd, Pb), Pb selenides and trace oxides and phosphates within closely spaced, bed-normal fractures (cleat) in several Middle Coal Measures coals of the anthracite region in the northwest of the coalfield. Gold was concentrated from these coals by ashing at 850 °C and maceration in hydrofluoric acid. The

grains are large (up to 100 × 100 × 20 µm), and tabular to flaky in form. Their thickness varies from 1 to >20 µm. The thinner grains have both flat surfaces smooth, whereas the thicker grains commonly have one surface showing the same colloform texture observed in the *in situ* grains on the cleat (*a* in the figure). In section, the grains consist of elongated gold microcrysts oriented normal to the flat surfaces. The thicker grains consist of more than one layer of microcrysts, each layer being about 1 µm thick (*b* in the figure). We interpret these relationships to indicate that the gold was precipitated on the cleat as a colloid, in association with the other exotic cleat minerals.

Subsequent semi-quantitative (inductively-coupled plasma) analyses of a suite of coals from the Middle Coal Measures showed significant but erratic levels of gold. These were confirmed by 23 assays (using the conventional lead button method on ashed coal samples), and showed gold values up to 4.42 g per tonne of ash, equivalent to 0.137 g per tonne in the coal. Significant variations were detected, both within correlated seams from sites about 40 km apart in the coalfield and from seams at different stratigraphic levels in the same site. The average gold content for the 23 samples of Middle Coal Measures coals in South Wales is 0.60 g per tonne of ash or 0.04 g per tonne of coal. A poor negative correlation between ash percentages and gold levels suggests that the gold is not of detrital origin, and agrees with the form of the gold grains indicating an origin as a colloidal precipitate within the cleat.

The origin of the gold is unclear. The Clarke value (the average value of a chemical element in the Earth's crust) for gold is not well constrained, but is thought to be around 0.005 g per tonne (ref. 3). The average gold content of these South Wales coals is about an order of magnitude greater than this. Swaine⁴ suggests that the average gold content of coal is <0.01 g per tonne, which suggests that the South Wales coals have an anomalously high value. Wedepohl³ reports that sedimentary rocks have

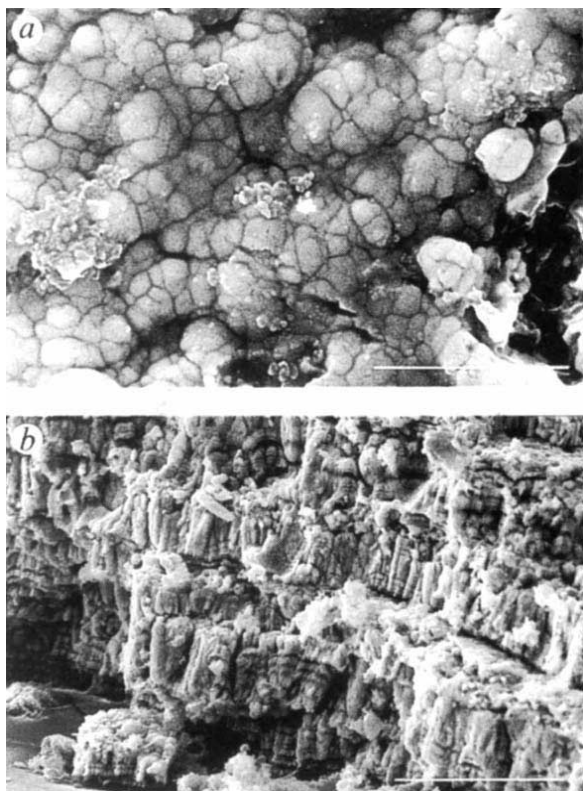
average gold contents around 0.003 g per tonne, somewhat below the Clarke, although there is a well known association of gold with organic-rich sediments giving levels two to three times the sedimentary rock average. Thus the average gold content of these South Wales coals seems high, although there is no evidence as yet that the amounts present are commercially significant.

There are no known gold deposits in South Wales that could have sourced the gold, nor are there any apparent igneous complexes. This does not mean that they do not exist, and the gold in South Wales coals may be an indication of a subsurface deposit. It is also possible that the source of the gold was the anthracite itself. The gold contents of terrestrial plants are poorly known, since most of the analyses come from areas with high bed-rock gold contents³. But Wedepohl³ suggests an average between 0.008 and 0.042 g per tonne of dry tissue. The association of the gold with an exotic carbonate, silicate and sulphide mineralogy on the coal cleat suggests that there is a common origin. A hydrothermal origin has been inferred⁵ for the cleat minerals involving deep fluid flow from the Variscan mountain belt of southwest Britain, guided northwards and upwards into the coalfield along a major linked thrust system.

Rod Gayer

David Rickard

*Department of Geology,
University of Wales, Cardiff CF1 3YE, UK*



Secondary electron scanning electron microscope images of a 100 × 100 × 20 µm gold grain from the Stwrin coal seam at East Pit opencast coal site, South Wales. The grain has been isolated by maceration in hydrofluoric acid from the coal ash. *a*, Surface of the gold grain, consisting of an assemblage of colloidal-sized particles. The underside of this grain is flat, suggesting that this specimen represents the free surface of the grain into the cleat opening. Scale bar, 2 µm. *b*, Cross-section through the gold grain, showing colloform banding in the gold layers. The banding represents textural variations displayed by varying responses in secondary electron imaging and may be associated with microcompositional changes in the gold. Scale bar, 10 µm.

1. van der Flier-Keller, E. *Econ. Geol.* **86**, 387–395 (1991).
2. Chyi, L. L. *Econ. Geol.* **77**, 1592–1597 (1982).
3. Wedepohl, K. H. (ed.) *Handbook of Geochemistry* (Springer, Heidelberg, 1978).
4. Swaine, D. J. *Trace Elements in Coal* (Butterworth, London, 1990).
5. Gayer, R. A. et al. *Proc. Osher Soc.* **7**, 380–384 (1991).

Whence Geminga?

SIR — Based on the measured proper motion of the star identified as the optical counterpart of Geminga (G'), Bignami *et al.*¹ and Gehrels and Chen² speculate that the Solar System may be within the supernova remnant that formed from the Geminga supernova event. But as neither the radial velocity nor the distance of the Geminga optical counterpart G' is known reliably, the actual space trajectory of Geminga is still poorly defined. I argue here that Geminga may instead have originated in the Orion association, a young association in which supernovae often occur, and in which there is evidence of supernova activity 300,000 years ago. In particular, Geminga may have caused the expanding supernova remnant 'Orion's cloak', which in turn may be linked to the youngest subgroup in Orion, Orion Id in the Trapezium, with an expansion age of around 300,000 years (refs 3, 4).

The parameters needed for the calcu-

lation of the Geminga space trajectory are the present location and distance from the Sun, the observed proper motion, and the age and radial velocity of Geminga. The values of these parameters, together with the uncertainties of those values, constitute the constraints on the Geminga space trajectory. The position of G' is $\alpha(1950) = 6\text{ h } 30\text{ min } 59\text{ s}$, $\delta(1950) = 17^\circ 48' 34''$; proper motions are $\mu_\alpha = 0.14'' \pm 0.05'' \text{ yr}^{-1}$ and $\mu_\delta = 0.10'' \pm 0.05'' \text{ yr}^{-1}$ (ref. 1). The spin-down age of Geminga is $\sim 340,000$ years^{1,2}, the uncertain distance 150–400 pc and the radial velocity is unmeasured. Using the transverse velocity range found for radio pulsars⁵, I selected an upper limit for both the space and radial velocities of Geminga of 400 km s^{-1} .

I calculated all Geminga space trajectories fitting these constraints using the programs described in ref. 6, in which solar motion is removed before calculating space trajectories. I compared the origin of each trajectory, identified as the Geminga location 340,000 years ago, to the positions of the Local Bubble, the entire Orion association and the Orion's cloak supernova remnant. Here, the Local Bubble is defined as all spatial points within 75 pc of the Sun, and the entire Orion association region as points in the interval $l=190^\circ\text{--}220^\circ$, $b=-30^\circ$ to -5° and distances 390–510 pc. The Orion's cloak supernova event is defined by points in the interval $l=194^\circ\text{--}210^\circ$, $b=-23^\circ$ to -7° , distances 390–510 pc. The volumes defined by the Local Bubble and Orion's cloak are comparable, but the volume defined by the entire Orion association is larger. Trajectories were calculated for every 1 km s^{-1} in velocity between -400 and 400 km s^{-1} , and every 10 pc in distance between 150 and 400 pc. I find that 2,610 trajectories originate in the entire Orion association with radial velocities between -340 and $+80 \text{ km s}^{-1}$; and 330 trajectories originate in the Local Bubble, with radial velocities between $+260$ and $+400 \text{ km s}^{-1}$. The ratio, R , of trajectories originating in the entire Orion association versus those in the Local Bubble is ~ 8 : 1,603 trajectories with radial velocities between -250 and $+80 \text{ km s}^{-1}$ originate in the Orion's cloak volume, and $R=5$.

Thus, the allowable class of trajectories contains significantly more trajectories originating in the Orion association than in the Local Bubble, even when equal volumes are defined. This conclusion is moderately robust. For a lower Geminga age of 300,000 years, $R=10$. Including

proper motion errors gives $R=0\text{--}10$. If a smaller radial velocity range, -300 to $+300 \text{ km s}^{-1}$, is assumed, $R\sim 25$. When the errors on the Geminga trajectory are reduced, it should be possible to identify exactly the supernova event caused by the Geminga explosion.

Priscilla C. Frisch

Department of Astronomy and
Astrophysics,
University of Chicago,
Chicago, Illinois 60637, USA

Age pigment structure

SIR — Earlier this year we reported the discovery of the general structure of one of the major autofluorescent substances that accumulates with age within the human retinal pigment epithelium (G. E. E. & M. R. Lasky *Nature* **361**, 724; 1993). Based largely on the results of fast-atom bombardment tandem mass spectrometry with collisionally activated decomposition analysis (FAB-CAD MS/MS) of hydrogenated and acetylated derivatives of the natural orange-emitting fluorophore, we proposed a general structure depicting two vitamin A molecules linked to the nitrogen of ethanolamine. The resulting quaternary amine contains 12 double bonds (consistent with hydrogenation experiments), five from each of the original retinaldehyde molecules plus one double bond from each of two proposed sequential Schiff base reactions.

Since publication, several investigators have suggested modifications of this structure that would be more consistent with the data published. First, the exact mass measurements of the underivatized native compound are inconsistent with the proposed structure. Assuming that the instrumentation was properly calibrated, the elemental composition derived from this data would dictate that there be one additional double bond in the molecule. Logical positions for an additional double bond would be between the two carbons of the ethanolamine moiety, or within the trimethylcyclohexenyl rings of one of the vitamin A moieties. Of these two, a ring desaturase reaction is not likely, therefore the former would be the favoured alternative. There are other alternatives: the ethanolamine head group could easily cyclize to form an additional ring. Vitamin A derivatives are also known to undergo 2 + 4 cyclodimerization reactions that would result in two fewer hydrogens. (I have searched for this structure using NMR and have been unable to detect it.) Alternatively, it has been suggested that the second vitamin A might be condensed to the carbon of the imine bond rather than the nitrogen. No new data exist that would favour any of these alternatives.

Until new studies are undertaken, no firm conclusion can be drawn. The general structure proposed remains consistent with the FAB-CAD MS/MS spectra.

A second observation is that the proposed structure was drawn as a zwitterion with the oxygen atom ionized. Several chemists have noted that the pK of this primary hydroxyl group would not favour its ionization. Alternatively, the balancing negative charge could be carried by an unspecified anionic counterion, or it could be dispersed throughout the extensive system of conjugated double bonds. Again, no new data exist that allow clarification of this issue.

Finally, as with any retinoid structure, an array of isomeric forms is possible, none of which can be distinguished using the techniques used in our work. In sum, as work proceeds on this new class of retinoid derivative, refinements of the proposed structure are likely to emerge.

Graig E. Eldred

Department of Ophthalmology,
University of Missouri,
Columbia, Missouri 65212, USA

Kinesin ATPase

SIR — Romberg and Vale¹ present data indirectly supporting the hypothesis that the affinity between kinesin and microtubules is lessened only after ATP is hydrolysed. If true, then the mechanism of force generation by kinesin is different from that of myosin, whose ATP-induced dissociation from actin precedes ATP hydrolysis².

This idea is not new: eight years ago Lasek and Brady³ reported that the non-hydrolysable ATP analogue AMP-PNP inhibits organelle transport by inducing tight binding between vesicles and microtubules. Although this earlier discovery was made in a biochemically ill-defined preparation, it nevertheless paved the way for the identification⁴ and partial purification⁵ of kinesin, which binds strongly to microtubules in the presence of AMP-PNP.

Romberg and Vale¹ should have cited the earlier work. But more important, before accepting the hypothesis as fact, we need direct evidence: the rate at which ATP dissociates the kinesin-microtubule complex needs to be measured using quenched flow or another rapid kinetic technique.

Jonathon Howard

Department of Physiology and Biophysics,
University of Washington,
Seattle, Washington 98195, USA

1. Bignami, G. F., Caraveo, P. A. & Mereghetti, S. *Nature* **361**, 704–706 (1993).
2. Gehrels, N. & Chen, W. *Nature* **361**, 706–707 (1993).
3. Cowie, L. L., Songaila, A. & York, D. G. *Astrophys. J.* **230**, 469–478 (1979).
4. Blaauw, A. *Rev. astron. Astrophys.* **2**, 213 (1964).
5. Lyne, A. G., Anderson, B. & Salter, M. J. *Mon. Not. R. astr. Soc.* **201**, 503–513 (1982).
6. Frisch, P. C. *Astrophys. J.* **407**, 198–209 (1993).

1. Romberg, L. & Vale, R. D. *Nature* **361**, 168–170 (1993).
2. Lynn, R. W. & Taylor, E. W. *Biochemistry* **10**, 4617–4624 (1971).
3. Lasek, R. J. & Brady, S. T. *Nature* **316**, 645–647 (1985).
4. Brady, S. T. *Nature* **317**, 73–75 (1985).
5. Vale, R. D. *et al. Cell* **42**, 39–50 (1985).

Instruments of darkness

Willem Hackmann

Radar at Sea: The Royal Navy in World War 2. By Derek Howse. Macmillan, UK: 1993. Pp. 383. £25.

DURING the early interwar years, news snippets appeared in the daily papers about mysterious experimental 'death rays' that could stop the engines of nearby motor cars, and in 1935 *The Daily Telegraph* reported on highly secret tests off the New Jersey coast of a 'ray' capable of detecting ships 50 miles away and aeroplanes at high altitudes. The story of radar (in particular airborne radar) has been told many times both as reminiscences by the pioneers — such as Robert Watson-Watt's *Three Steps to Victory* (1957) and Taffy Bowen's *Radar Days* (1987) — and as 'official' or 'semi-official' histories, but this is the first book specifically on the British Royal Navy's involvement.

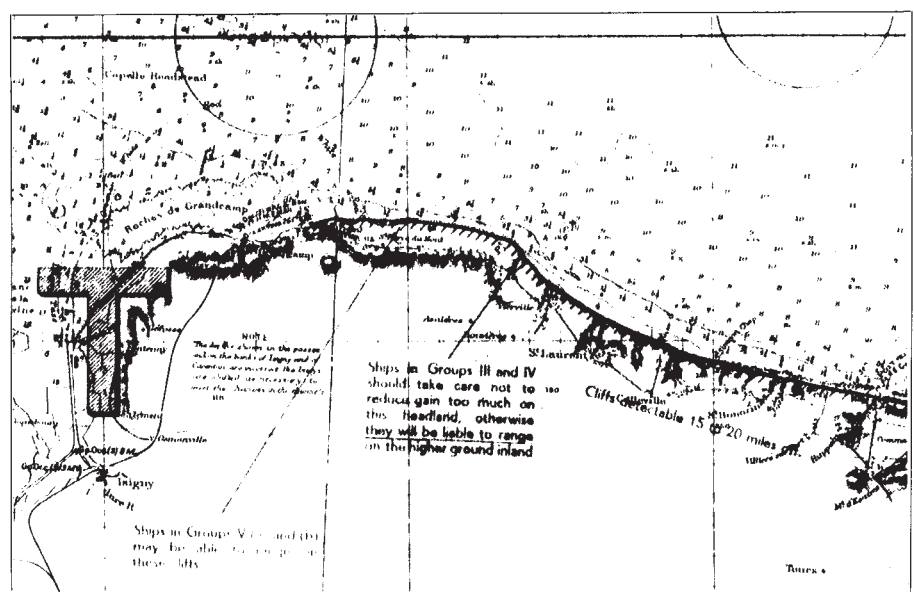
This lacuna in the history of radar was keenly felt by Admiralty scientists and naval officers concerned with the development of naval radar during the Second World War, and they decided to do something about it. Being of a practical disposition, they established the Naval Radar Trust, which commissioned this book and provided technical assistance. Similar concerns had already motivated another group of Admiralty scientists to commission a history of the Royal Navy's involvement in asdic (sonar to those born after 1943), which was published by Her Majesty's Stationery Office in 1984. Both projects have resulted in the preservation of valuable documents and recorded interviews. Derek Howse, who was approached by the Naval Radar Trust, has first-hand experience of the sharp end of radar, serving at sea in the Royal Navy throughout the war and later specializing in navigation. He is also an experienced naval historian (albeit of an earlier period) and was a curator at the National Maritime Museum in Greenwich after taking early retirement from the Navy.

In March 1935, the only instrument available on Royal Navy ships for detecting aircraft was a pair of binoculars of magnification $\times 7$. Shortly after the Munich crisis, the battleship *Rodney* and the cruiser *Sheffield* could detect aircraft 60 miles away with apparatus that had not even been on the drawing board four years before. This book chronicles the evolution and influence of naval radar during the nine momentous years from late 1936, when the first experimental set (type 79X) was fitted for trials in the elderly minesweeper *Salisbury*, to 1945, when the cruiser *Norfolk* was fitted with the latest radars (type 274 for gunnery and types 277, 293 and 181B as warning sets) and

other electronic gear for action information and target indication, such as the plan position indicator, skiatron and automatic plotting table. The ship's bridge had traditionally been the place from which the captain fought the battle, but by the end of the Second World War technology had speeded up warfare so dramatically that the bridge was rapidly superseded by the (electronic) action information centre in the bowels of the ship.

reader is assisted by an excellent index, detailed appendices (including one on naval airborne radar by David Brown) and a table of the Royal Navy's radar sets covering the period 1935–45.

In late 1943 the Germans started attacks on the British fleet using aircraft-launched radio-controlled glider-bombs, assaults that caused consternation. A naval scientist happened to switch on his electric razor during one of these attacks in the Bay of Biscay, and to the amazement of the ship's crew the bomb began to gyrate about the sky and eventually gave chase to the very aircraft that had launched it. An order was issued by the Admiralty Signal Establishment that when ships came under these attacks, all able-bodied men who had electric razors should rush on deck, plug their instruments into the



Engaging the enemy — Detail from the radar chart for the assault on the Normandy coast, published 15 March 1944.

The author tells this complex story by weaving together the diverse strands of technical development, naval policy, scientific bureaucracy and operations at sea, in chronological sequence. To cope with this approach, he has broken up the narrative with numerous subheadings. Specific radar developments are thus placed primarily within the framework of naval operations, such as the sinking of the *Hood* and *Bismarck* in 1941, the German development in anti-ship guided weapons in 1943 and the 'radar war' preceding the Normandy landings in 1944. This approach ensures the smooth flow of the narrative, but it can also be irksome when one tries to follow certain specific radar developments, which are often broken off and then resumed some pages further on under another subheading. So the serious reader will have to do a fair amount of leafing forwards and backwards. There is no easy answer to this problem, however, and in this case the

nearest power point and wave them wildly at the offending missile(s). The effectiveness of this countermeasure is not recorded, but it must have boosted morale until electronic jammers were developed.

The author's analysis (reinforced by the book's format) appears to be that wartime radar development was essentially 'event-driven': that is, it developed as a response (or countermeasure) to enemy action. That this is largely the case is borne out by the parallel development of sonar in which, for instance, the glider-bomb episode is mirrored by the Navy's response to the German acoustic torpedo. Both sonar and radar began life as 'add-on' substitutes for optical range-finding — as superior gun sights — and both evolved into sophisticated weapon systems. This history, concentrating as it does on the development of the various types of Royal Navy radar (the fruits of, in the author's words, "a race between operational thought and technical achievement") is

useful as far as it goes. But it would perhaps have been even more interesting if the author had gone beyond describing the immediate technical responses to particular situations and tackled some of the broader issues, such as the context of radar development in the Royal Navy's general scientific research strategy, the involvement of industry in this process, and the factors that influenced radar development by Britain's principal allies and enemies. This may well have been outside Howse's brief, however, and the Naval Radar Trust is in any case intending to publish another volume purely on the technical development of naval radar.

Howse has managed to bring the Royal Navy's involvement in radar to life, not least in the novel 'tailpieces' at the end of the chapters, where he presents amusing anecdotes (such as the electric razor story). His book is a fitting tribute to all the radar scientists, naval officers and men who devoted their energies to develop these 'electronic eyes', sometimes at the cost of their lives. □

Willem Hackmann is at the Museum of the History of Science, Old Ashmolean Building, Broad Street, Oxford OX1 3AZ, UK.

Other worlds

Constantine Pagonis

The Quantum Theory of Motion: An Account of the de Broglie-Bohm Causal Interpretation of Quantum Mechanics. By Peter R. Holland. Cambridge University Press: 1993. Pp. 598. \$120, £70.

IN 1952, David Bohm did what many thought had been proved impossible: picking up from where Louis de Broglie had left off 25 years earlier, he formulated a "hidden variables" theory of quantum mechanics. In other words, he came up with a "quantum theory of motion", making quantum mechanics a sort of statistical mechanics arising out of our ignorance of the detailed trajectory dynamics.

Bohm's 1952 theory is a (non relativistic) deterministic theory that captures all the predictions of quantum mechanics. Roughly speaking, the picture is of particles whose trajectories are determined by their initial positions and a new type of physical field, the ψ field. The particle positions at any time together with the ψ field determine the values that a measurement would yield at any time for all dynamical quantities associated with the particles. The fact that most quantum predictions are probabilistic reflects our ignorance of the (uncontrollable) initial positions of the particles rather than any irreducibly stochastic behaviour of the system.

In Bohm's theory, measurement interactions do not have special status compared with other types of interactions (there is no 'collapse' of the wavefunction), and complementary experimental situations do not have the effect of making some properties of microscopic systems defined and others undefined: all properties have precise values at all times. That is not to say that Bohm's theory is a classical theory of quantum phenomena: it has a number of novel features that are highly nonclassical (it is a nonlocal and contextual theory, for example). But, however nonclassical, all its features have a solid physical basis.

Reversing Bohr's famous dictum, Bohm might have said: "There is a quantum world". For, unlike the standard interpretation of quantum mechanics, this was a theory of quantum mechanical processes in space and time, not just a theory about experimental results. Physicists could now imagine a quantum world, substituting physics for dogmatic constraints on what questions could be asked. The days of the standard interpretation were surely numbered.

Alas, talk of the mysterious 'collapse' of the wavefunction and of wave-particle duality was not banished to fringe publications exploring the 'connections' between Eastern mysticism and modern physics, but is still found in almost all physics courses taught today. Why this should be is unclear. After reading Peter Holland's book, however, it becomes apparent that the answer must lie in the sociology of twentieth-century physics, for it does not lie in the physics itself.

The Quantum Theory of Motion is a tour de force. After a first chapter on the problem of interpreting quantum mechanics, Holland sets the scene with a review of classical Hamilton-Jacobi theory, which he generalizes, making Bohm's quantum theory of motion and classical mechanics special cases. He then presents the elements of Bohm's theory and proceeds to apply it to all the standard one-body and many-body quantum mechanical problems, developing the theory to account for spin and discussing its extension into the relativistic domain. There is a very useful chapter on the analysis of measurements in quantum mechanics according to Bohm's theory, and special attention is paid throughout to the classical limit of the theory as well as to its novel features.

I strongly recommend *The Quantum Theory of Motion* to the physics community for its clear and thorough presentation of a theory deserving much wider attention. □

Constantine Pagonis is in the Department of History and Philosophy of Science, University of Cambridge, Free School Lane, Cambridge CB2 3RH, UK.

Thoughts and impressions

Sara J. Shettleworth

Through Our Eyes Only? The Search for Animal Consciousness. By Marian Stamp Dawkins. W. H. Freeman: 1993. Pp. 192. £14.95, \$19.95.

Do nonhuman animals have conscious experiences? This question has already been the subject of three books by Donald Griffin (the latest is *Animal Minds*, Chicago University Press, 1992, which was reviewed in *Nature* 360, 118 (1992)). It often arises in philosophical treatments of consciousness and is increasingly discussed by cognitive psychologists. What Marian Dawkins contributes to this debate is not new ideas or examples so much as a short, clear statement, for lay people or beginning students, of the issues involved.

In her previous book, *Animal Suffering* (Chapman and Hall, 1980, now unfortunately out of print), Dawkins displayed a talent for unpacking issues in animal behaviour that are normally surrounded by a great deal of woolly thinking. *Through Our Eyes Only?* continues this tradition. Her aim is to air both sides of the debate about whether animals are conscious by presenting evidence for animal consciousness, arguments against that evidence, more evidence that satisfies stricter criteria and so on. So anyone not reading the book from beginning to end (not an arduous task) is likely to come away with a dangerously incomplete impression.

Dawkins's thesis goes like this: subtleties of perception control of behaviour by complex stimuli and feats of learning and memory are sometimes taken to mean that animals must be conscious. But complex behaviour does not necessarily imply complex processing, nor does complex processing necessarily imply consciousness. For one to conclude that complex cognitive processes underlie apparently complex decision-making, one must eliminate the Clever Hans effect (unintentional cueing by experimenters), have a clear statement of the chance of the observed behaviour under other hypotheses, and — perhaps most difficult — eliminate the possibility that the behaviour is controlled by some combination of simple rules of thumb. Accomplishing all this still does not address the questions of consciousness directly.

Nevertheless, Dawkins claims that in a few cases, such as counting by rats and deception by primates, animals seem to be consciously weighing alternative courses of action. Here, as elsewhere, she explains examples from experimental animal

psychology, ethology and behavioural ecology fully enough for the general reader to understand what was done and why certain conclusions were drawn. The problem with such examples, of course, is that they stand or fall on the ingenuity of the sceptics' alternative hypotheses.

In the penultimate chapter, Dawkins returns to the theme of her earlier book to discuss whether animals experience emotions as well as thoughts like ours. Here she comes down clearly on the side of those who believe that they do, arguing that the behavioural evidence makes this conclusion far simpler than any other.

Why should anyone care whether animals are conscious? I must confess to falling firmly among those whom Dawkins classifies as sceptics. We dismiss calls to study animal consciousness because to do so amounts to trying to answer an unanswerable question about the quality of private experience. How and why animals do what they do can be analysed effectively without assuming that consciousness is involved in behaviour. But I have been given pause by Dawkins's point that to dismiss the question of animal consciousness is in effect to place consciousness outside biology. If consciousness has evolved, it must have some effect on behaviour that we can eventually detect. A consistent darwinian cannot accept that humans or other animals are conscious while saying that the effects of consciousness are undetectable in behaviour, at least not without accepting the suggestion that consciousness is no more than an epiphenomenon of certain kinds of brain activity.

Although Dawkins casts doubts on entrenched positions of sceptics and believers alike, the book's illustrations (described in her preface as "visual images that are exactly right") give her away as being far from a sceptic herself. What else are we to conclude when a raccoon, five-fingered paws clasped, stares out at the reader from the page facing a chapter titled "Balance of Evidence"? Each chapter is headed not only by a photograph (most of them less provocative than the anthropomorphic raccoon) but also by a passage from Fred Hoyle's *The Black Cloud*. The black cloud, it may be recalled, turned out against all probabilities to be a conscious being, capable of communicating with humans. I remain unconvinced that the same is true of nonhuman animals, but anyone wanting a thoughtful guide to thinking about the possibility of consciousness in monkeys, bees or black clouds could do no better than to read Dawkins's book. □

Sara J. Shettleworth is in the Departments of Psychology and Zoology, University of Toronto, 100 St George St, Toronto, Ontario M5S 1A1, Canada.



SCALES of the bark of the lemon-scented gum tree *Eucalyptus citriodora* (left) and spines of a red *Pereskia* cactus from Brazil (right). Taken from *Bark: The Formation, Characteristics, and Uses of Bark Around the World*, with text by G. T. and A. E. Prance and photographs by K. B. Sandved. Timber Press, \$49.95, £37.50.

Heat treatment

Michael L. May

The Hot-Blooded Insects. By Bernd Heinrich. Harvard University Press/Springer (UK): 1993. Pp. 601. \$75, £78.

SOME moths fly using muscles at temperatures near freezing point; other insects such as ants are most active at around 50 °C. This and similar evidence, Bernd Heinrich argues, shows that high body temperature and high energy metabolism are not necessarily advantageous and that temperature and the precision of its regulation are, within broad limits, evolutionary options influenced by the morphology and behavioural ecology of each insect, as well as the physics of its environment. The argument may be pushed a bit beyond its limits; Heinrich himself cites evidence that flight at low temperature requires compromises in the optimal design of flight muscle. Nevertheless, his careful and logical questioning of what drives the evolution of thermoregulation is one of the real strengths of *The Hot-Blooded Insects*. The book is certainly 'adaptationist', but its arguments for the adaptedness of insect thermoregulatory strategies are convincing and largely independent of any phylogenetic analyses that remain to be done.

In the same vein, Heinrich also challenges the idea that insect wings evolved initially as lateral outgrowths for thermoregulation. He cogently argues that flight must have preceded effective regulation of body temperature, although he perhaps

too easily dismisses the quantitative success of this model in bridging the gap between rudimentary lobes and large, aerodynamic surfaces.

The book's other great strength lies in its thorough coverage of a great array of literature (there are at least several hundred citations, some dating back to 1742). There is, as one might expect from such a prolific contributor to the field, considerable emphasis on his own research. Anyone with an emerging interest in the subject could hardly find a better introduction to what has been done.

The organization is largely taxonomic (a chapter on moths, one on butterflies, one on orthopterans (grasshoppers, crickets, locusts) and so on). It works fairly well, although it does lead to some repetition. A chapter that departs from the taxonomic theme, "Cold Jumpers", deals with the problem of how insects that do *not* regulate temperature generate rapid bursts of locomotory power. Unfortunately, the discussion becomes so enmeshed in complex, incompletely understood mechanical detail that its relation to thermoregulation tends to be lost. The parallel drawn between catch mechanisms of jumpers and elastic storage in flight is tenuous, based as it is on the now discredited 'click-mechanism' of flies proposed by E. Boettinger and E. Furshpan. In fact, despite Heinrich's repeated emphasis on the importance of flight in driving the evolution of thermoregulation, his treatment of the mechanics and physiology of flight is diffuse and often not very up to date.

The seemingly gratuitous summary chapter is also disappointing. The

proposition that optimal body temperatures should be evolutionarily conservative may be true but appears to contradict much of the argument made so effectively in earlier chapters; and the support presented is cursory at best. The last section on future directions is disingenuous because its list of "redundant" studies could discourage worthwhile research on the ecology of thermoregulation, and the research 'wish list' is too narrowly focused. There is still a crying need to understand how temperature and its control really affect insects' lives. But a valid point is made: although it is no longer news that insects regulate their body temperature and that this ability affects many other aspects of their biology, some of the physiological mechanisms by which this is done are not so well understood as they might seem.

The Hot-Blooded Insects is an outstand-

ing source of information, and can be read with profit and satisfaction by the professional biologist and interested amateur alike. Professionals may find the frequent parenthetical definitions of familiar terms distracting while lay people may wish for more guidance through a few physiological thickets. True initiates in the arcana of insect thermoregulation may be somewhat less pleased, partly because Heinrich presents little new information and batters some pet theory of nearly everyone in the field. Some of the horses he flogs are moribund, if not long dead. On the other hand, he raises many salutary issues on a subject that still has much to contribute to the study of insects and adaptation. □

Michael L. May is in the Department of Entomology, Rutgers University, PO Box 231, New Brunswick, New Jersey 08903, USA.

Accidents and complexity

H. J. Melosh

Solar System Evolution: A New Perspective. By Stuart Ross Taylor. *Cambridge University Press: 1993. Pp. 307. £35, \$49.95.*

THIRTY years ago, scientific study of the Solar System was an unpopular backwater of astronomy, shunned by observers who set their sights much farther afield. Speculation about the Solar System's origin was restricted to a few eccentric elder scientists who were not constrained by any real facts about the composition or geological history of the planets. Since then, however, dozens of spacecraft have flown through the Solar System, sending back high-quality images and other data on every planet except Pluto. Samples of rock and soil have been brought back from the Moon, and vastly improved analytical methods have yielded greater insight into the history of meteorites. Planetary science has now blossomed into an active interdisciplinary research area that borrows techniques and ideas from both astronomy and geology.

Many early workers hoped that these findings would quickly reveal a simple comprehensive picture of the Solar System in which such fundamental properties as the densities and compositions of the planets would show readily explainable regularities, and the structure of the Solar System could be seen as an inevitable consequence of the collapse of some kind of protostellar cloud of gas and dust. Instead, the data have revealed a Solar System that is wildly diverse. It seems that stochastic processes have dominated every stage of the Solar System's evolution. Giant impacts of hundreds of planet-

sized protoplanets randomly determined the orbits and rotational states of the few survivors. The compositions of the giant planets were established by a hasty, last-minute competition between accretion onto protoplanetary cores and dissipation of the solar nebula by violent outbursts of the young Sun. Even the meteorites, presumed to be slightly altered samples of the most ancient solid bodies in the Solar System, show a numbing complexity that has stubbornly resisted broad generalizations. Although the full implications of the new catastrophic view of Solar System formation have yet to be spelled out, the outlines of the new-world picture are becoming clearer.

Reducing the vast amount of data to manageable proportions is a monumental task. It is easy to get sidetracked on knotty but fascinating problems and lose sight of the overall picture. Stuart Ross Taylor, a geochemist by profession, has taken a broad view of Solar System evolution, and his most recent book is a clear distillation of his work. Although he naturally inclines to geo- or cosmo-chemical arguments, he has not ignored constraints from physics, astronomy and geology, making this the first coherent account of the new view of Solar System history. Although most of the individual facts in the book can be found in one or another of the massive tomes in the University of Arizona Press's Space Science Series (to which Taylor makes repeated references), this book offers a readable and concise introduction.

After a brief historical and philosophical discussion of planetary formation, Taylor describes nebular accretion models, putting the formation of our Solar

System into the context of the Universe and the Galaxy. The discussion of nebular collapse is necessarily vague and will probably be rapidly revised as radio-astronomical techniques continue to resolve star formation in nearby molecular clouds. Moving from theory to evidence, he devotes his longest chapter to meteorites, offering a solid, no-nonsense separation of facts from nonfacts and of plausible theories from wishful thinking. Not everyone will agree with his conclusions, but he does discuss the main issues in an unbiased way. In his next chapter, he covers impacts, a testimony to the importance recently attached to impact cratering. Much of the discussion centres on the lunar cratering record, but most of the other principal topics are touched on. He then gives a planet-by-planet description of the main bodies in the Solar System, with emphasis on the geochemical properties of the terrestrial planets and asteroids, although the giant planets are also briefly described. Finally, he discusses rings and satellites while debunking the myth that planetary satellite systems are miniature Solar Systems. As is to be expected of a book fresh from the front lines of scientific research, many of the issues are not yet satisfactorily resolved, such as the origins of chondrules, the age of planetary ring systems, and the nature of heat sources in the early solar nebula; but Taylor gives a balanced discussion of what is known about each of these problems. He ends his book with a philosophical consideration of the complexity of the Solar System.

This is not a book for beginners or those on undergraduate courses. Without some acquaintance with geochemistry and isotopic dating methods readers will be quickly lost. Topics are presented roughly in order of their appearance in the evolving Solar System, a format which often presumes that the reader has knowledge of facts and ideas long before they are introduced in detail. The discussion of meteorites therefore begins with calcium-aluminum inclusions, proceeds to chondrules and only later addresses meteorites themselves. This approach demands quite a bit of repetitive discussion of the same topic in different contexts, and often breaks up the presentation. For graduate courses, however, where students have been previously acquainted with planetary science, the book should serve as an excellent reference: it succinctly summarizes evidence and arguments. But its main audience will be professional Earth and planetary scientists or knowledgeable amateurs who want a readable introduction to the latest ideas on the origin and evolution of the Solar System. □

H. J. Melosh is in the Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA.

Transcriptional antitermination

Jack Greenblatt, Justin R. Nodwell & Stephen W. Mason

Antiterminator proteins control gene expression by recognizing control signals near the promoter and preventing transcriptional termination which would otherwise occur at sites that may be a long way downstream. The N protein of bacteriophage λ recognizes a sequence in the nascent RNA, and modifies RNA polymerase by catalysing the formation of a stable ribonucleo-protein complex on its surface, whereas the λ Q protein recognizes a sequence in the DNA. These mechanisms of antitermination in λ provide models for analysing antitermination in viruses such as HIV-1 and in eukaryotic genes.

BOTH in prokaryotes and in eukaryotes, gene expression is often controlled by altering the efficiency with which transcription terminates at various sites (reviewed in refs 1–4). Antitermination mechanisms may be divided according to whether they are processive or not. In non-processive mechanisms, a regulatory protein or translating bacterial ribosome interacts with the nascent RNA transcript, controlling the activity of a nearby transcriptional terminator or ‘attenuator’^{5,6}. By contrast, in processive mechanisms the RNA polymerase is modified so that it transcribes through multiple transcriptional terminators which may be located far downstream from the promoter. Here we consider only the latter class of events. The best characterized examples of processive mechanisms are found in the lambdoid bacteriophages, where recent studies provide instructive examples for considering antitermination during the transcription of eukaryotic and viral genes. We shall focus on transcriptional antitermination in phage λ and in the human immunodeficiency virus, HIV-1.

Antitermination in phage λ

After infection of *Escherichia coli* by phage λ , the phage N protein prevents termination at multiple sites in both early operons^{1,4} (Fig. 1a). This allows *E. coli* RNA polymerase to transcribe more efficiently the *O* and *P* genes, required for replicating the phage DNA, and the *Q* gene, located further downstream in the pR early operon. The Q protein, in turn, prevents termination during transcription of the late pR operon, causing more efficient expression of the late genes involved in bacteriophage morphogenesis and in lysis of the infected cell. This cascade of antitermination factors regulates the timing of λ gene transcription during a cycle of lytic infection. The early

operons are targets for N-protein action because they contain N-utilization (*nut*) sites between their promoters and the first terminator in the operon, whereas the late operon is regulated by Q because it contains a Q-utilization (*qut*) site within its promoter (reviewed in refs 1 and 4). Some λ terminators, including tL1 and tR1, require the *E. coli* termination factor Rho to function (reviewed in ref. 7), but N and Q prevent termination at any terminator located downstream of a λ *nut* site or *qut* site, whether Rho-dependent or not.

Antitermination by the N protein on phage λ

The current model for antitermination by the λ N protein is shown in Fig. 2; a closely related system regulates the synthesis of ribosomal RNA in *E. coli*. A key feature of these systems is that the control signals are located in the nascent RNA, and are connected to the elongating RNA polymerase by RNA looping and multiple protein–protein interactions. The elongation factors are bound to the RNA polymerase as it transcribes DNA downstream from the control signals.

Assembly of the N-modified elongation complex. The isolation of *E. coli* strains (*nus* mutants) that do not support antitermination by N led to the identification of three host factors, NusA, NusB and NusE, which modulate termination and antitermination in *E. coli* (reviewed in refs 1, 4); a fourth, NusG, was recently identified by the ability of a *nusG* mutation to suppress the effects of the *nusA1* mutation on antitermination by N (ref. 11). Interestingly, the *nusE* gene product is ribosomal protein S10 (ref. 8).

NusA and N by themselves can cause antitermination *in vitro* at Rho-dependent and -independent terminators, provided the terminator is sufficiently close to the *nut* site^{9,10} (see Fig. 2a). In

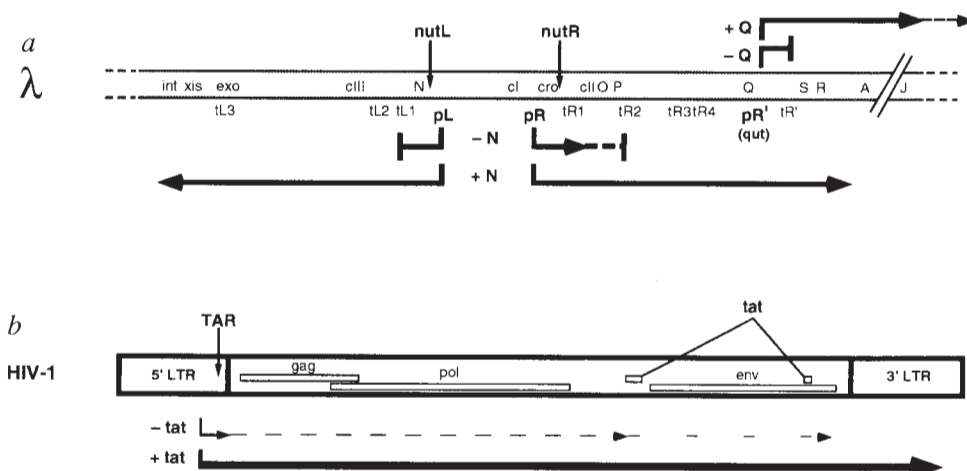


FIG. 1 a, Partial genetic and transcriptional map of bacteriophage λ , indicating the effects of the λ N and Q proteins. b, Partial genetic and transcriptional map of the HIV-1 proviral genome, showing the effect of *tat* on transcription.

in vivo, however, antitermination is highly processive, and N can prevent termination at sites located as much as 5–10 kilobases (kb) downstream from the *nutL* or *nutR* sites¹ (Fig. 1a). *In vitro*, N will produce processive antitermination in reactions containing purified NusA, NusB, NusG and S10 (refs 10, 12) (Fig. 2b). In these reactions N and all four host factors are stably and stoichiometrically bound to the elongating transcription complex, and formations of a stable complex requires a *nut* site in the template DNA^{13–15}. NusA associates with RNA polymerase soon after the latter releases its initiation subunit σ^{70} (refs 13, 16), whereas N can bind any time after the polymerase has moved past the *nut* site^{15,17}. The observation that all five factors are needed to form a stable complex correlates well with the fact that all five are also necessary for processive antitermination^{10,14,18}.

Assembly of the termination-resistant form of RNA polymerase induced by N depends on multiple protein–protein interactions within the ternary transcription complex. Interactions between N and NusA¹⁹ and between NusA and RNA polymerase¹⁶ form the basis of the minimal antitermination system involving N and NusA^{9,10,17} (see Fig. 2a). Like NusA, NusG

and S10 bind to RNA polymerase^{12,14}, but their association is weak and unstable in the absence of a *nut* site and the other factors¹⁴. As shown in Fig. 2b, N and NusB associate indirectly with RNA polymerase through their interactions with NusA and S10, respectively^{19,20}. All of the protein–protein interactions in the N-modified transcription complex are weak (dissociation constant $K_d \geq 5 \times 10^{-8}$ M), which may explain why assembly of the complex is highly cooperative^{14,18}.

The *nut* site is made of RNA. The *nut* sites required for antitermination by N have two sub-elements, *boxA* and *boxB* (reviewed in refs 1, 4; Fig. 3b). The λ *boxA* element is highly conserved in the antiterminator elements of various lambdoid phages, such as phages 21 and P22 (refs 1, 4, 21–23), but the λ *boxB* is not²³. The function of λ N is specific to *nut* sites containing the λ *boxB*, as opposed to those of phages 21 and P22, suggesting that N recognizes *boxB* (reviewed in refs 1, 24). N contains an arginine-rich motif near its amino terminus which is responsible for recognizing the appropriate *nut* site, and this region is believed to be an RNA-binding domain²⁴.

Many lines of evidence have shown that the functional *nut* site is in the nascent RNA, rather than the DNA^{13,17,18,25,26}. More recently, each sub-element of the site has been shown to bind a specific subset of the elongation factors (Fig. 3b). N itself binds to the *boxB* RNA in an electrophoretic gel mobility shift assay²⁷, whereas NusA binds to the N-*nut* site complex, presumably by virtue of its interaction with N (J.R.N. and J.G., unpublished data). The resulting ternary complex weakly protects *boxB* RNA, but not *boxA* RNA, from ribonuclease digestion during transcription *in vitro*¹⁸.

The seven ribosomal RNA (*rrn*) operons of *E. coli*, which encode the 5S, 16S and 23S rRNAs, as well as some transfer RNAs, each contain an antiterminator element. These elements, also referred to as *boxA*, have sequences very similar to the *boxA* elements of the λ *nut* sites^{1,21,22} (see Fig. 3a, b), and are sufficient to suppress termination when placed between a promoter and terminator (reviewed in ref. 22). The resemblance between *rrn* elements and the λ *boxA* suggested that the host factors involved in antitermination by N might also mediate *rrn* antitermination. Indeed, the *nusB5* mutation impairs synthesis of rRNA *in vivo* (reviewed in refs 1, 4, 22), and antitermination during *rrn* transcription with *E. coli* extracts *in vitro* requires NusB protein²⁸. NusB and S10 form a heterodimer²⁰ which binds *rrn boxA* RNA, and *rrn boxA* mutations which affect binding of the NusB–S10 complex³⁰ have a commensurate effect on antitermination²⁹. Unlike the *rrn boxA* elements, the λ *boxA* elements do not bind the NusB–S10 complex or function as antiterminators in the absence of *boxB*^{1,30}. The λ *boxA* thus behaves as a defective form of *rrn boxA* whose ability to bind the NusB–S10 complex and induce antitermination has come to depend on N and *boxB*. Only when N and all four host factors are present is the λ *boxA* sequence protected from RNase during transcription *in vitro*¹⁸.

RNA looping. How does RNA polymerase maintain a termination-resistant form far downstream of the antiterminator sequence? One possibility is that an RNA loop initially connects the λ *nut* site RNA to the polymerase³¹ (Fig. 2), so that polymerase is still associated with the *nut* site when it arrives at a downstream terminator. This idea has recently received substantial support. Mutations in β -subunit of the polymerase that interfere with antitermination by N also interfere with *nut* site protection during transcription *in vitro*¹⁸, indicating that the surface of the polymerase aligns the elongation factors so that they can bind the *nut* site RNA. Moreover, N can engage the elongating polymerase after it has passed the *nut* site even when the *nut* site DNA has been removed from the template by a restriction enzyme¹⁷. This observation is only easy to understand in the context of an RNA looping model (Fig. 2). The RNA would initially serve as a 'tether' to assist formation of the complex by holding N in the vicinity of the polymerase¹⁸. Similarly, NusB, S10 and *rrn boxA* RNA may well use RNA looping to form a

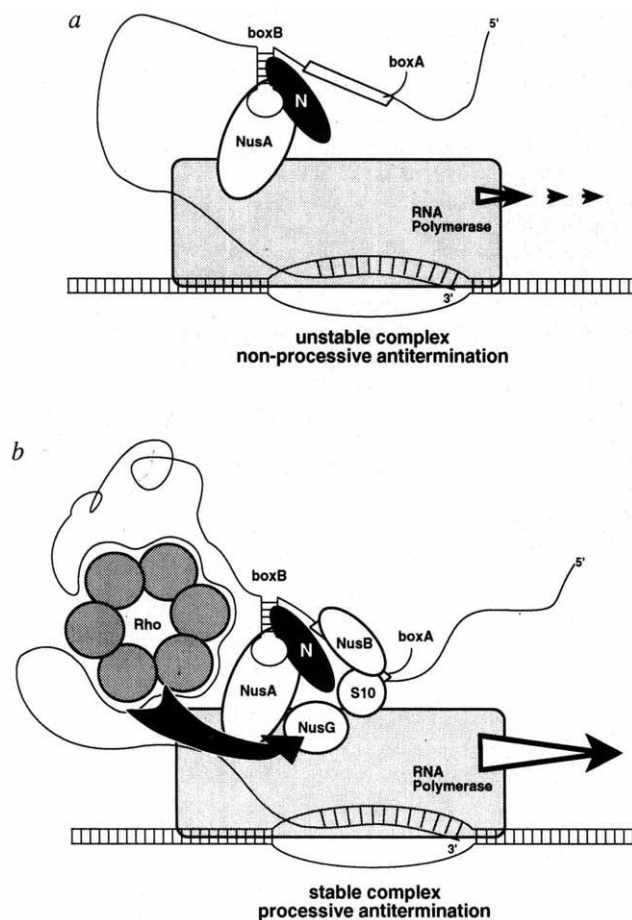


FIG. 2 Antitermination by the λ N protein involves RNA looping and multiple protein–protein interactions. a, Non-processive antitermination involves an unstable complex containing N and NusA. We have suggested that N prevents termination, in this case, because it suppresses pausing by *E. coli* RNA polymerase¹⁰. b, Processive antitermination involves a stable complex containing N, NusA, NusB, NusG and ribosomal protein S10. We have hypothesized that the indicated interaction between NusG and Rho factor prevents Rho from releasing the nascent RNA transcript³⁴. A similar mechanism may prevent release of RNA chain during transcription of the *E. coli rrn* operons³⁰.

ribonucleoprotein complex on the surface of RNA polymerase that also includes NusA and NusG¹². Such a complex would lack the λ N protein, but otherwise be similar to the one shown in Fig. 2b. Once a stable complex involving all the host factors has formed on the surface of the polymerase, as shown in Fig. 2b, the RNA loop is presumably dispensable. Indeed, the primary transcripts of the *rrn* operons must be processed to form mature rRNAs and tRNAs²² and the pL RNA of λ is cleaved between *nutL* and tL1 *in vivo*²².

Mechanisms of antitermination. When transcription terminates, the RNA polymerase pauses, and then releases the RNA transcript (reviewed in ref. 2). An antiterminator protein could, in principle, inhibit either of these steps. In the case of a Rho-dependent terminator, Rho factor moves along untranslated nascent RNA by virtue of its RNA-dependent ATPase activity and eventually unwinds the RNA-DNA hybrid at the active site of the RNA polymerase (reviewed in ref. 7). Whether transcription is terminated or not seems to depend on whether the Rho factor moves sufficiently fast to 'catch up' with the polymerase³³, and so N could prevent termination by increasing the polymerase's speed. In fact, N, NusA and *boxB* RNA do suppress pausing by the polymerase at sites in the Rho-dependent λ terminator tR1 (ref. 10). This antipause activity may explain why N and NusA can prevent termination *in vitro* at Rho-dependent or Rho-independent terminators without NusB, NusG or S10 (refs 9, 10). It is still unclear how N prevents pausing by RNA poly-

merase, but one possibility is that the N protein or *boxB* RNA occupies a site on the polymerase that a portion of the nascent RNA must contact to induce pausing. Otherwise, N may accelerate the polymerase simply by stabilizing a conformation of the polymerase which does not allow pausing.

N may also indirectly inhibit Rho-dependent release of the nascent transcript. Affinity chromatography with immobilized NusG has shown that NusG binds directly to Rho factor³⁴. NusG also binds weakly to RNA polymerase¹² and is important for Rho factor activity *in vivo* and *in vitro*^{34,35}, suggesting that it acts as a bridge between the polymerase and Rho to induce termination of transcription. The Rho-NusG interaction, however, may also be important for antitermination by N, because the *rho026* mutation makes termination by Rho less dependent on NusG and also makes both antitermination by N and the binding of NusG to Rho temperature-sensitive³⁴. N may therefore prevent termination by stabilizing the association between NusG and the transcription complex^{12,14} in such a way that the interaction between NusG and Rho prevents the latter from reaching the RNA-DNA hybrid at the active site of the polymerase (Fig. 2b). In contrast, the loosely bound NusG molecule in an unmodified transcription complex somehow helps Rho to terminate transcription.

The *rrn boxA* antitermination system only affects Rho-dependent terminators and not strong Rho-independent terminators like those at the ends of the *rrn* operons³⁶. As termination at all terminators is apparently inhibited when RNA polymerase pausing is suppressed by proteins such as λ N and Q^{4,10,37}, the *rrn boxA* system may not affect pausing by RNA polymerase. Elongation complexes synthesizing rRNA *in vitro* contain NusG¹², and an anti-release mechanism involving the sequestration of Rho factor by NusG, like that shown in Fig. 2b, for λ , could explain why *rrn boxA* only affects Rho-dependent terminators.

Antitermination by the Q protein of Phage λ

Q prevents pausing by RNA polymerase. RNA polymerase molecules initiating transcription at the λ late promoter pR' pause for several minutes 16–17 nucleotides downstream from the initiation site^{4,38} (Figs 3d and 4), and molecules leaving this site terminate transcription at the simple terminator tR' in the absence of Q protein. Q protein recognizes the paused transcription complex as a substrate and accelerates the polymerase out of the pause site and through tR' into the late gene region (reviewed in ref. 4). Q can prevent pausing, not only at +16, but also at the pause sites of the λ Rho-dependent terminator tR1 when that terminator is artificially fused to pR' (ref. 37). The ability of Q to inhibit pausing may explain its ability to suppress termination at all terminators, whether Rho-dependent or not. The host factors required for Q activity seem to be much simpler than those required by N. Q can act alone on the stalled polymerase complex at +16, but the addition of NusA helps antitermination and is essential at some terminators³⁷. Although multiple host factors are needed to maintain the processivity of N-mediated antitermination^{10,12}, there is no indication that additional host factors are needed to maintain the processivity of Q-mediated antitermination.

Q is a DNA-binding protein. In striking contrast to the *nut* sites, which are in the nascent RNA, the *qut* site needed for antitermination by Q lies in the DNA between bases -26 to +18 of the late promoter pR' (reviewed in ref. 4) (Fig. 3d). The transcribed region of *qut* downstream of +1 acts as a pause signal, as point mutations at +2 and +6 prevent both pausing by the polymerase at +16 and antitermination by Q (ref. 38). The mutation at +2 abolishes Q activity even if RNA polymerase is stalled at +16 by nucleotide deprivation³⁹, suggesting that the nucleotides downstream of +1 may not only stall the polymerase, but also cause the transcription complex to adopt a conformation allowing modification by Q.

Mutations at -13 and -15 within the pR' promoter prevent

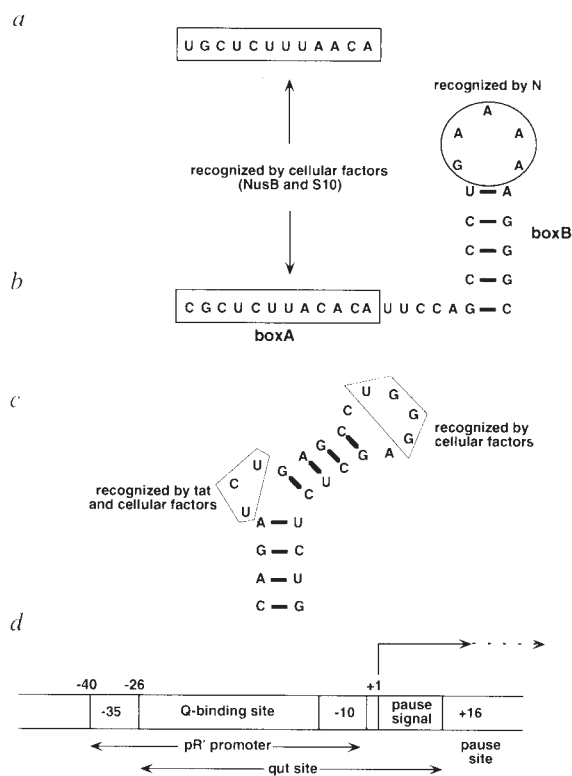


FIG. 3 Interactions of host and viral antitermination factors with various antitermination elements. a, Recognition of *rrn boxA* RNA by the *E. coli* antitermination factors, NusB and ribosomal protein S10 (ref. 30). b, λ *nut* sites are made of RNA and consist of two functional elements, *boxA* and *boxB*. *BoxA* is recognized by NusB and S10, and *boxB* is recognized by the λ N protein²⁷. c, Recognition of the HIV-1 TAR element RNA by human host factors and viral tat protein. Tat recognizes the bulge sequence UCU in the stem of TAR (reviewed in ref. 43). d, The λ *qut* site overlaps the λ late promoter pR' and is recognized by the λ Q protein upstream of the transcriptional initiation site³⁹. Sequences in the *qut* site downstream of -1 cause RNA polymerase to pause and adopt a conformation that is receptive to 'modification' by Q.

the binding of Q to the DNA between -10 and -30 and block antitermination³⁹. Although Q can bind to this sequence in the absence of RNA polymerase, it binds to the same region when the transcription complex is stalled at +16 (Fig. 4). If NusA is present, Q then induces changes in the contacts between RNA polymerase and the promoter, especially between -2 and +6. It is possible that these changes are maintained during chain elongation and are important for the action of Q.

How does Q modify RNA polymerase? How Q prevents termination at sites far downstream from pR' is unknown³⁹. For antiterminator proteins that bind to RNA, like the λ N protein, the association of the control signal and antiterminator protein with the transcription complex can be maintained by RNA looping³¹ (Fig. 2). For a DNA-binding antiterminator protein like Q, the analogous possibility is DNA looping (model A in Fig. 4), but the single Q molecule bound to each *qut* site could only contact one RNA polymerase molecule transcribing the λ late operon.

Another possibility is that Q is released from the DNA and travels with RNA polymerase when it leaves the pause site at +16 (model B in Fig. 4). Because Q cannot affect the polymerase without a *qut* site, it must trap a particular conformation of the polymerase induced by the *qut* site or Q at +16 which forces Q to remain bound to the polymerase during elongation. Otherwise, Q may induce a conformational change in the polymerase at +16 which is maintained during elongation without further direct contact by Q (model C in Fig. 4). Experiments that distin-

guish among the models shown in Fig. 4 will be crucial for understanding how Q acts.

Related control systems in eukaryotes. Transcriptional anti-termination is now known to regulate the expression of many eukaryotic genes. For example, uninduced *Drosophila hsp70* genes carry an arrested molecule of RNA polymerase II about 25 base pairs downstream from the transcriptional initiation site (reviewed in ref. 40). The appearance of this arrested polymerase molecule depends on sites in the *hsp70* promoter which bind the GAGA transcription factor and on sequences downstream from the initiation site which presumably cause the polymerase to pause⁴⁰. In response to heat shock, the heat shock factor (HSF) binds to other sites in the promoter and releases the paused polymerase, a situation remarkably similar to transcriptional regulation by Q protein (Fig. 4). Many other insect and mammalian genes, including *c-myc*^{41,42}, are similarly regulated (reviewed in ref. 3), and processivity must be very important for the transcription of extremely long genes like dystrophin. It is easy to imagine that activator proteins such as HSF contact RNA polymerase II or one of its associated factors in order to modify the elongation properties of the enzyme.

Antitermination by the tat protein of HIV-1

Efficient synthesis of HIV-1 proteins requires the virally encoded protein tat (Fig. 1b) (reviewed in refs 43, 44). There is now considerable evidence that tat both activates transcriptional initiation from the promoter in the 5' LTR of HIV-1 (refs 45, 46) and enhances the processivity of elongating transcription complexes in a manner reminiscent of the λ N protein^{24,45,47,48}.

Much of the evidence that tat is a transcriptional antiterminator comes from nuclear run-on transcription experiments^{45,47,48}. In the absence of tat, the transcripts hybridize to promoter-proximal DNA fragments, but not to fragments further downstream from the promoter. In the presence of tat, long RNAs accumulate which hybridize to DNA fragments from anywhere in the transcription unit, indicating that tat causes RNA polymerase II to read through transcriptional blocks. Discrete sites of transcription termination in HIV-1 DNA have not been identified, but most transcripts synthesized in the absence of tat end upstream of the *gag* gene⁴⁸ (Fig. 1b). When tat is added to *in vitro* transcription reactions in HELA nuclear extracts, it dramatically increases the yield of RNA from the promoter of the HIV-1 LTR^{49,50}, again largely if not entirely by increasing the processivity of transcription complexes.

The TAR site recognized by tat is made of RNA. Tat antitermination depends on an element called TAR located downstream from the transcription start in the HIV-1 LTR (Fig. 1b) (reviewed in ref. 43). The minimal TAR site between bases +19 and +42 of the HIV-1 transcript is located within a region of dyad symmetry which extends from +1 to +59. A transcript of this region can form an RNA hairpin with a three nucleotide bulge and a six nucleotide loop (Fig. 3c), both of which are required for tat function. Mutations in the stem of the hairpin which disrupt its base-pairing potential also impair transactivation by tat, unless compensatory mutations which restore base-pairing in the stem are also present, showing that the sequence of the stem is less important than its secondary structure. These and other observations demonstrate that the functional form of the TAR element, like the *nut* sites of bacteriophage λ , is made of RNA rather than DNA.

Electrophoretic mobility shift and RNA footprinting experiments show that tat recognizes the three-nucleotide bulge in TAR RNA (reviewed in ref. 43). The interaction between the bulge and tat seems functionally important as changes in the bulge sequence which impair tat binding have a corresponding effect on transactivation by tat. It has been proposed that TAR is recognized by a unique 'arginine fork' motif in the basic region of tat⁵¹, which has a very similar composition to the arginine-rich domain used by N protein to recognize the λ *nut* sites^{24,52-54}. The striking structural and functional homologies

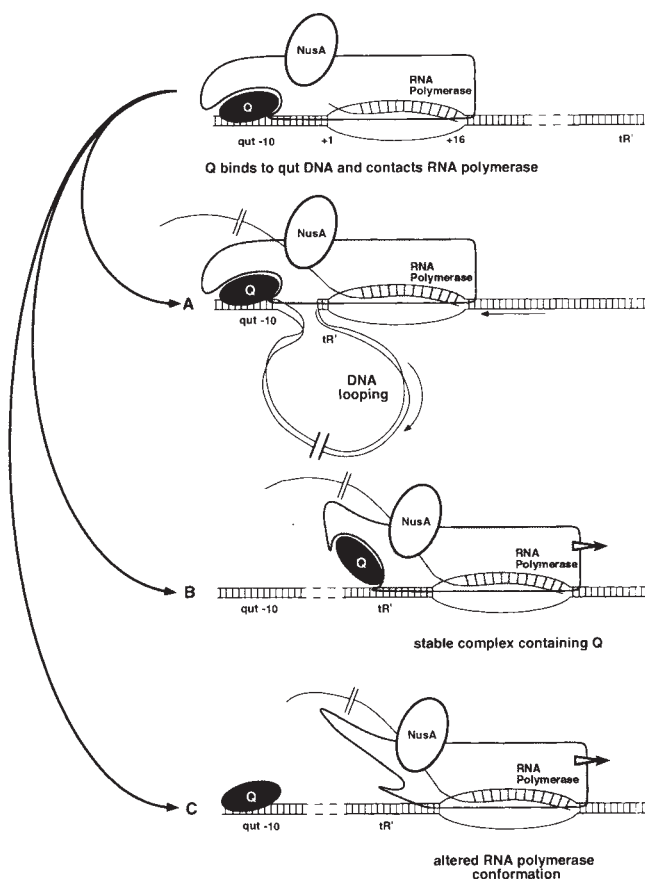


FIG. 4 Possible models for antitermination by the λ Q protein. Initially, Q binds to *qut* site DNA and interacts with RNA polymerase that has paused at +16 (ref. 39). Model A: DNA looping might enable Q to maintain simultaneous contact with RNA polymerase and the *qut* site. Model B: Q might be released from the *qut* site and travel in stable association with the RNA polymerase. Model C: Q might modify the conformation or structure of RNA polymerase in such a way that the alteration is maintained during transcription despite the lack of further direct contact by Q.

between tat and N suggest that tat, like N, uses RNA looping to make contact with the transcription complex^{31,44} (Figs 2 and 5).

Human host factors for tat. At least two types of human proteins seem to affect transactivation by tat. These interact with TAR and tat, respectively. Mutational studies of the TAR loop (reviewed in ref. 43) and functional assays *in vitro* indicate that several HeLa cell proteins bind to the TAR loop and affect transactivation by tat^{55,56}. Despite the apparent biological relevance of these proteins, fusions between tat and other RNA-binding proteins such as the rev protein of HIV-1 (ref. 57) or the coat protein of bacteriophage R17 (ref. 53) activate transcription

when the TAR sequence is replaced with the sequence recognized by the RNA-binding part of the fusion, indicating that the only essential role of TAR is to interact with tat and deliver it to the transcription complex by RNA looping (Fig. 5). The human proteins that bind to TAR RNA may thus be important only insofar as they modulate tat binding.

Other human proteins probably bind directly to tat. Deletion of tat residues 2–20 or 22–34 from a tat-R17 coat protein fusion abolishes transactivation both through TAR RNA and through the R17 operator, indicating that these parts of tat may interact with host factors essential for transactivation⁵³. A 36K human protein which improves transactivation by tat in mouse cells was isolated by affinity chromatography on columns containing immobilized tat⁵⁸. Another tat-binding protein, TBP-1, was cloned by probing a human expression library with purified tat protein⁵⁹, and inhibits transactivation by tat when overexpressed in transfected cells. More recently, the complementary DNA for MSS1, a human protein 42% identical to TBP-1, was isolated⁶⁰. Cotransfection of cDNAs encoding MSS1 and tat into P19 embryonal carcinoma cells greatly stimulates the activity of a reporter construct driven by the HIV-1 LTR. These results indicate that MSS1 and TBP-1 may regulate transactivation by tat.

Models for tat function. Superficially, there are striking parallels between antitermination activities of tat and the λ N protein. Both use similar RNA-binding domains to recognize control signals in nascent transcripts, and both bind host factors that are important for their function (NusA in the case of N and possibly MSS1 in the case of tat). Additional host factors bind to both the *nut* site RNA and the TAR RNA (Fig. 3b, c), suggesting that tat, like N (Fig. 2), could assemble into a ribonucleoprotein complex with host factors and TAR RNA on the surface of RNA polymerase (Fig. 5).

Some of these host elongation factors may be general factors required for initiation by RNA polymerase II. The human initiation factor TFIIF (RAP30/74) binds to RNA polymerase II and recruits it to a pre-initiation complex containing the initiation factors TFIID and TFIIB (reviewed in refs 61, 62). As TFIIF also increases the speed of chain elongation and suppresses pausing at certain sites⁶³, it behaves in some ways like a combination of N and NusA¹⁰. Recently, TFIIF was found to be important for antitermination by tat *in vitro*, and processivity in the absence of tat can be improved by adding excess TFIIF⁵⁰. Tat may thus prevent termination by aiding the interaction between TFIIF and RNA polymerase II during elongation, and proteins that bind to tat, like MSS1 and TBP-1 (refs 59, 60), might aid or inhibit this interaction. The initiation factors TFIIE, TFIIH and TFIIJ associate with the initiation complex after TFIIF and RNA polymerase II (reviewed in ref. 62), and may also remain associated with the polymerase during elongation (Fig. 5). Any of them could be a direct target of tat. As TFIIH is a protein kinase^{64,65}, the interaction between tat and the initiation or elongation complex may even induce covalent modification of the latter, stabilizing the binding of TFIIF and permanently altering the elongation properties of RNA polymerase II so that there is no further need for contact between the transcription complex and tat.

There are also major differences between the mechanisms by which tat and N act. For example, tat fused to the DNA-binding domains of the transcription factors JUN and GAL4 can activate transcription independently of TAR, if an appropriate DNA recognition element is inserted upstream from the TATA box in the HIV-1 promoter^{46,66}. Transcriptional activation by GAL4-tat⁴⁶ may explain why tat has been found to affect transcriptional initiation as well as elongation⁴⁵. Proteins that activate initiation like human Spl, herpes simplex virus VP16, and adenovirus E1A may bind to coactivators that interact with TBP, the DNA-binding subunit of the TATA box factor TFIID, as well as binding TBP and the general initiation factor TFIIB directly (for recent reviews, see refs 61, 62, 67). The effects of tat on initiation may therefore involve interactions with general

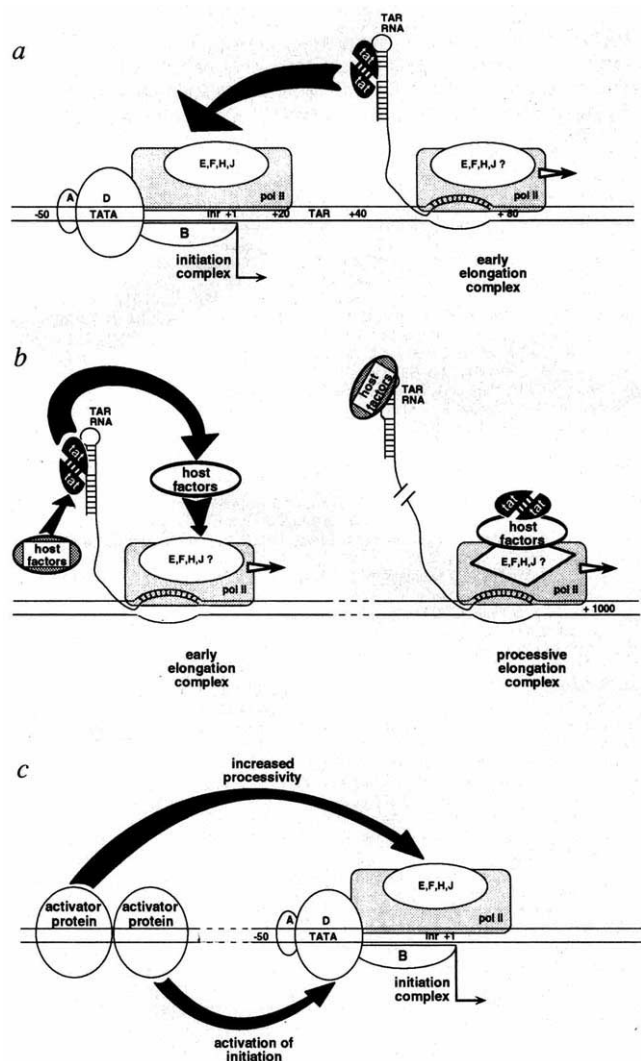


FIG. 5 a, Activation of transcription initiation by tat involves an RNA enhancer mechanism⁴⁴. Tat protein that is bound to TAR RNA has to 'reach back' and make contact with the next initiation complex at HIV-1 promoter to stimulate initiation of transcription. TFIIA, B, D, E, F, H and J are general factors required for initiation at most promoters by RNA polymerase II (reviewed in ref. 62). b, Transcriptional antitermination by tat involves an RNA looping mechanism. We suggest that tat is eventually released from the TAR site⁵³ and that host factors may have a role in this process⁵⁶. Tat would then interact with host elongation factors that associate with RNA polymerase II. This interaction may stabilize the association with RNA polymerase II of a host general factor (TFIIE, TFIIF, TFIIH or TFIIJ) or may cause modification of a host general factor, thereby increasing the processivity of chain elongation by RNA polymerase II. c, Transactivation by DNA binding transcription factors may also involve a contact that increases the processivity of chain elongation by RNA polymerase II.

initiation factors and/or coactivators. As an RNA-binding protein, tat would have to 'reach back' from its binding site in an elongation complex by RNA looping to stimulate assembly of the subsequent initiation complex at the promoter⁴⁴ (Fig. 5a).

The way in which tat regulates transcription may be universal. When tat functions through TAR, it has a dramatic effect on transcription activated by the weak chimaeric activator GAL4-Spl, but only a modest effect when transcription is activated by the strong chimaeric activators GAL4-VP16 and GAL-E1A⁴⁶. A simple interpretation of this data is that strong activators such as VP16, E1A and tat, but not weak activators such as Spl, increase the processivity of transcription as well as the assembly of active initiation complexes. Tat would thus differ from other strong activator proteins only in binding to RNA rather than DNA. One prediction of this hypothesis, which was recently confirmed⁶⁸, is that a strong, DNA-binding activator protein, such as VP16, should also work when bound to RNA. Another prediction, which has still to be tested, is that strong transactivators such as VP16 should increase the processivity of chain elongation. These observations can readily be understood by

postulating that each strong activator can contact, directly or indirectly, a general factor regulating chain elongations, as well as one or more factors required only for initiation (Fig. 5c). Contact with the elongation factor would lead to a long-lasting increase in the processivity of the elongation complex. The requirement that a strong activator contact at least two general transcription factors could then explain why the action of most DNA-binding activators is highly cooperative (for review, see ref. 61). The model shown in Fig. 5, which combines concepts derived from λ N antiterminator system (RNA looping; association of several elongation factors with RNA polymerase; Fig. 2) with both putative host factors and general factors regulating initiation and elongation, could explain antitermination in many eukaryotic systems. □

Banting and Best Department of Medical Research and Department of Molecular and Medical Genetics, University of Toronto, 112 College Street, Toronto, Ontario, Canada M5G 1L6. J.R.N.'s present address: Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138, USA.

- Friedman, D. I. In *The Bacteriophages* Vol. 2 (ed. Calendar, R.) 263-318 (Plenum, New York, 1988).
- Yager, T. D. & Von Hippel, P. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*. Vol. 2 (ed. Neidhardt, N. C.) 1241-1275 (American Society for Microbiology, Washington DC, 1987).
- Spencer, C. A. & Groudine, M. *Oncogene* **5**, 777-785 (1990).
- Roberts, J. W. *Cell* **52**, 5-6 (1988).
- Amster-Choder, O. & Wright, A. *Science* **257**, 1395-1398 (1992).
- Yanofsky, C. *J. biol. Chem.* **263**, 609-612 (1988).
- Richardson, J. P. *Cell* **64**, 1047-1049 (1991).
- Friedman, D. I., Schauer, A. T., Baumann, M. R., Baron, L. S. & Adhya, S. L. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1115-1118 (1981).
- Whalen, W., Ghosh, B. & Das, A. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2494-2498 (1988).
- Mason, S. W., Li, J. & Greenblatt, J. *J. biol. Chem.* **267**, 19418-19426 (1992).
- Sullivan, S. L., Ward, D. F. & Gottesman, M. E. *J. Bact.* **174**, 1339-1344 (1992).
- Li, J., Horwitz, R., McCracken, S. & Greenblatt, J. *J. biol. Chem.* **267**, 6012-6019 (1992).
- Horwitz, R. J., Li, J. & Greenblatt, J. *Cell* **51**, 631-641 (1987).
- Mason, S. W. & Greenblatt, J. *Genes Dev.* **5**, 1504-1512 (1991).
- Barik, S., Ghosh, B., Whalen, W., Lazinski, D. & Das, A. *Cell* **50**, 885-899 (1987).
- Greenblatt, J. & Li, J. *Cell* **24**, 421-428 (1981).
- Whalen, W. A. & Das, A. *New Biol.* **2**, 975-991 (1990).
- Nodwell, J. R. & Greenblatt, J. *Genes Dev.* **5**, 2141-2151 (1991).
- Greenblatt, J. & Li, J. *J. molec. Biol.* **147**, 11-23 (1981).
- Mason, S. W., Li, J. & Greenblatt, J. *J. molec. Biol.* **223**, 55-66 (1992).
- Li, S. C., Squires, C. L. & Squires, C. *Cell* **38**, 851-860 (1984).
- Morgan, E. A. *J. Bact.* **168**, 1-5 (1986).
- Franklin, N. C. *J. molec. Biol.* **181**, 85-91 (1985).
- Lazinski, D., Grzadzinska, E. & Das, A. *Cell* **59**, 207-218 (1989).
- Olsen, E. R., Tomich, C.-S. C. & Friedman, D. I. *J. molec. Biol.* **180**, 1053-1063 (1984).
- Warren, F. & Das, A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3612-3616 (1984).
- Das, A. *J. Bact.* **174**, 6711-6716 (1992).
- Squires, C. L., Greenblatt, J., Li, J., Condon, C. & Squires, C. L. *Proc. natn. Acad. Sci. U.S.A.* **90**, 970-974 (1993).
- Berg, K. L., Squires, C. L. & Squires, C. *J. molec. Biol.* **209**, 345-358 (1989).
- Nodwell, J. R. & Greenblatt, J. *Cell* **72**, 261-268 (1993).
- Greenblatt, J. *Can. J. biochem. cell. Biol.* **62**, 79-88 (1984).
- Lozeron, H. A., Aneviski, P. J. & Apioiron, D. *J. molec. Biol.* **109**, 359-365 (1977).
- Jin, D. J., Burgess, R. R., Richardson, J. P. & Gross, C. A. *Proc. natn. Acad. Sci. U.S.A.* **89**, 1453-1457 (1992).
- Li, J., Mason, S. W. & Greenblatt, J. *Genes Dev.* **7**, 161-172 (1993).
- Sullivan, S. L. & Gottesman, M. E. *Cell* **68**, 989-994 (1992).
- Albrechtson, B., Squires, C. L., Li, S. & Squires, C. *J. molec. Biol.* **213**, 123-134 (1990).
- Yang, X. & Roberts, J. *Proc. natn. Acad. Sci. U.S.A.* **86**, 5301-5305 (1989).
- Kainz, M. & Roberts, J. *Science* **255**, 838-841 (1992).
- Yarnell, W. S. & Roberts, J. W. *Cell* **69**, 1181-1189 (1992).
- Lee, H.-S., Kraus, K. W., Wolfner, M. F. & Lis, J. T. *Genes Dev.* **6**, 284-295 (1992).
- Krumm, A., Meulia, T., Brunvand, M. & Groudine, M. *Genes Dev.* **6**, 2201-2213 (1992).
- Strohl, L. J. & Eick, D. *EMBO J.* **11**, 3307-3314 (1992).
- Cullen, B. R. *Cell* **63**, 655-657 (1990).
- Sharp, P. A. & Marciniak, R. A. *Cell* **59**, 229-230 (1989).
- Lasplia, M. F., Rice, A. P. & Mathews, M. B. *Cell* **59**, 283-292 (1989).
- Southgate, C. D. & Green, M. R. *Genes Dev.* **5**, 2496-2507 (1991).
- Kao, S., Calman, A. F., Luciw, P. A. & Peterlin, B. M. *Nature* **330**, 489-493 (1987).
- Feinberg, M. B., Baltimore, D. & Frankel, A. D. *Proc. natn. Acad. Sci. U.S.A.* **88**, 4045-4049 (1991).
- Marciniak, R. A. & Sharp, P. A. *EMBO J.* **10**, 4189-4196 (1991).
- Kato, H. et al. *Genes Dev.* **6**, 655-666 (1992).
- Calnan, B. J., Tidor, B., Biancalana, S., Hudson, D. & Frankel, A. D. *Science* **252**, 1167-1171 (1991).
- Weeks, D. M., Ampe, C., Schultz, S. C., Steitz, T. A. & Crothers, D. M. *Science* **249**, 1281-1285 (1990).
- Selby, M. J. & Peterlin, B. M. *Cell* **62**, 769-776 (1990).
- Calnan, B. J., Biancalana, S., Hudson, D. & Frankel, A. D. *Genes Dev.* **5**, 201-210 (1991).
- Marciniak, R. A., Garcia-Blanco, M. A. & Sharp, P. A. *Proc. natn. Acad. Sci. U.S.A.* **87**, 3624-3618 (1990).
- Sheline, C. T., Milocco, L. H. & Jones, K. A. *Genes Dev.* **5**, 2508-2520 (1991).
- Southgate, C. D., Zapp, M. L. & Green, M. R. *Nature* **345**, 640-642 (1990).
- Desai, K., Loewenstein, P. M. & Green, M. *Proc. natn. Acad. Sci. U.S.A.* **88**, 8875-8879 (1991).
- Nelbock, P., Dillon, P. J., Perkins, A. & Rosen, C. A. *Science* **248**, 1650-1653 (1990).
- Shibuya, H. et al. *Nature* **357**, 700-702 (1992).
- Greenblatt, J. *Cell* **66**, 1067-1070 (1991).
- Zawel, L. & Reinberg, D. *Curr. Opin. Cell Biol.* **4**, 488-495 (1992).
- Bengal, E., Osvaldo, F., Krauskopf, A., Reinberg, D. & Aloni, Y. *Molec. cell. Biol.* **11**, 1195-1206 (1991).
- Lu, H., Zawel, L., Fisher, L., Egly, J.-M. & Reinberg, D. *Nature* **358**, 641-645 (1992).
- Serizawa, H., Conaway, R. C. & Conaway, J. W. *Proc. natn. Acad. Sci. U.S.A.* **89**, 7476-7480 (1992).
- Berkhout, B., Gatignol, A., Robson, A. B. & Jeang, K.-T. *Cell* **62**, 757-767 (1990).
- Pugh, B. F. & Tjian, R. *J. biol. Chem.* **267**, 679-682 (1992).
- Tiley, L. S., Madore, S. J., Malim, M. H. & Cullen, B. R. *Genes Dev.* **6**, 2077-2087 (1992).

ACKNOWLEDGEMENTS. We thank J. Ingles and J. Segall for their constructive criticism of our manuscript and to J. Ingles for suggesting how to improve the artwork for our figures.

Extending the Vostok ice-core record of palaeoclimate to the penultimate glacial period

J. Jouzel^{*†}, N. I. Barkov[‡], J. M. Barnola[§], M. Bender^{||}, J. Chappellaz[§], C. Genthon[§], V. M. Kotlyakov[¶], V. Lipenkov[‡], C. Lorius[§], J. R. Petit^{§*}, D. Raynaud[§], G. Raisbeck[#], C. Ritz[§], T. Sowers^{||}, M. Stievenard^{*}, F. Yiou[#] & P. Yiou^{*}

^{*} Laboratoire de Modélisation du Climat et de l'Environnement, CEA/DSM CE Saclay, 91191, Gif sur Yvette Cedex, France

[†] Arctic and Antarctic Research Institute, Beringa Street 38, 199226, St Petersburg, Russia

[§] Laboratoire de Glaciologie et Géophysique de l'Environnement, CNRS, BP96, 38402, Saint Martin d'Hères Cedex, France

^{||} Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island 02882-1197, USA

[¶] Institute of Geography, Staromonetny, per 29, 109017, Moscow, Russia

[#] Centre de Spectrométrie Nucléaire et de Spectrométrie de Masse, 91405, Orsay, France

The ice-core record of local temperature, dust accumulation and air composition at Vostok station, Antarctica, now extends back to the penultimate glacial period (~140–200 kyr ago) and the end of the preceding interglacial. This yields a new glaciological timescale for the whole record, which is consistent with ocean records. Temperatures at Vostok appear to have been more uniformly cold in the penultimate glacial period than in the most recent one. Concentrations of CO₂ and CH₄ correlate well with temperature throughout the record.

IN 1985 Lorius *et al.*¹ presented the first ice-core climate record spanning a full glacial–interglacial cycle. The record, which went back to the end of the penultimate glacial period about 150,000 years ago (150 kyr before present, BP) was based on ice drilled at Vostok, the Russian station in central East Antarctica (78° 28' S and 106° 48' E, mean annual temperature –55 °C, elevation 3,490 m). The Vostok records show that East Antarctica was colder and drier during glacial periods than during the Holocene^{1–4}, demonstrate that the large-scale atmospheric circulation was more vigorous during glacial times^{5,6}, support evidence from deep-sea sediment studies for orbital forcing of Pleistocene climate^{2,7} and reveal direct correlations of carbon dioxide and methane concentrations with temperature. The last point suggests that variations in greenhouse gas concentrations have contributed to glacial–interglacial changes in climate^{8–11}. Most of this work was based on the study of ice from the surface to 2,083 m depth in the 3Γ core.

Here we extend the record of Vostok geochemistry to the bottom of a new core, 4Γ, drilled down to 2,546 m. We include details of deuterium and dust concentrations measured in the ice and three parameters measured in the air bubbles entrapped in the ice: CO₂, CH₄ and the oxygen isotope ratio (given as δ¹⁸O) of O₂. Each parameter contains climate-related information. The new results (Table 1) are plotted against depth in Fig. 1, along with data from the bottom part of core 3Γ. The shallowest depth plotted, 1,500 m, corresponds to the end of the last interglacial period (LIG) and is dated at about 110 kyr BP. In addition, we will use information on changes in accumulation rate at Vostok contained in a new ¹⁰Be profile also measured down to 2,546 m. Details on experimental procedures may be found elsewhere^{2,6,8,9,12,13}. The depth profiles of CO₂, CH₄ and δ¹⁸O in O₂ will undoubtedly evolve as additional samples are analysed.

We first describe a new timescale for the entire Vostok core.

After discussing the reconstructed Vostok temperature record, we test our chronology by comparing various Vostok climate records with climate records from deep-sea sediments and, where relevant, curve of insolation. We then discuss the climate information contained in our new data set which covers the entire penultimate glacial period and extends through a part of the next to last interglacial down to ~220 kyr BP.

Extending the glaciological dating approach

To interpret the Vostok data, we need a chronology for the ice, expressed as a curve of age against depth. We derive this chronology from the following four assumptions. First, the accumulation rate at Vostok has varied in the past as the derivative of the water vapour saturation pressure with respect to temperature at the level where the precipitation forms; that is, above the inversion layer^{1,14}. Second, the accumulation (*A*) upstream of Vostok, where ice found at depth originates, increases with distance upstream according to $A(x)/A(x=0) = 1.00 + 0.65(x/X)$, where *x* is distance upstream of Vostok and *X* is distance of site Dome B upstream of Vostok (320 km). This equation invokes the observation that the Holocene as reflected by the δD or δ¹⁸O records starts at a depth of ~300 m at Vostok and ~500 m at Dome B¹⁵. Third, ice at a depth of 1,534 m has an age of 110 kyr BP^{16,17}. Finally, the strain-induced thinning of annual layers with depth is accurately described by the two-dimensional glaciological model of Ritz¹⁸.

With these assumptions, we calculate the accumulation rate at Vostok today to be 1.98 g cm⁻² yr⁻¹, 13% lower than the value of 2.3 g cm⁻² yr⁻¹ (ref. 1) based on estimates over the past 10 years, but in good agreement with the value of 2.00 ± 0.4 g cm⁻² yr⁻¹ estimated over the past 170 years using the Tambora eruption as a marker¹⁹. Back to 110 kyr BP, the biggest discrepancy with the chronology of Lorius *et al.*¹ is 3.5 kyr (Fig. 2). The age of the bottom of the core is 220 kyr BP. We refer to our age–depth curve as the 'extended glaciological timescale' (EGT). The uncertainty in the age of the bottom of the 4Γ core (2,546 m), largely dominated by the uncertainty in the current accumulation rate¹⁸, is estimated to be ±20 kyr.

[†] J.J. is also at the Laboratoire de Glaciologie et Géophysique de l'Environnement. M.B. is on sabbatical leave at the Centre des Faibles Radioactivités, CNRS-CEA 91198, Gif-sur-Yvette Cedex, France.

Raisbeck *et al.*¹³ obtained ^{10}Be concentrations in Vostok and have subsequently analysed the 3 Γ and 4 Γ cores in higher resolution (Raisbeck *et al.*, manuscript in preparation). We have used these data to calculate a curve of accumulation rate against depth for Vostok, assuming that the ^{10}Be deposition rate is constant and the same as the average value observed over the top 1,534 m (as above, we assume an age of 110 kyr BP for this depth). Complications that might introduce errors in such an estimate of accumulation are discussed in refs 20 and 21. Combining this accumulation rate with the flow model used above, we can calculate a ^{10}Be chronology which differs by no more than ~ 5 kyr from the EGT (Fig. 2). Although we believe this concordance tends to support the overall correctness of the EGT, it does not preclude significant differences in detail, particularly over periods when there were important climatic changes. For

example, this ^{10}Be chronology gives a duration for the Holocene $\sim 30\%$ shorter than that given by the EGT.

The isotope temperature record

Precipitation is isotopically enriched with respect to atmospheric water vapour, and the vapour remaining in a given air mass is depleted at each condensation step both in δD and $\delta^{18}\text{O}$. Simple models^{22,23} predict that δ should vary linearly with temperature, T , in mid- and high latitudes. This predicted dependence is well documented from observations of present-day Antarctic precipitation^{24,25}. We use the modern δ - T relationship to translate the δ change observed going back in time, at one given location, into a curve of temperature against age. This approach is supported by results from the general climate model studies^{26,27} and independent glaciological estimates of glacial-interglacial temperature changes²⁸.

Temperature reconstructed from the isotopic profile is plotted against EGT age in Fig. 3 (curve 3b). We calculate the change from the present temperature using the equation $\Delta T_a = (1^\circ\text{C per } 9\text{‰}) \times (\delta\text{D}_{\text{ice}} - 8\delta^{18}\text{O}_{\text{sw}})$. T_a is the temperature immediately above the inversion layer where Antarctic precipitation forms¹⁴. The first term on the right is the inverse of the dependence of δD on temperature observed for this sector of East Antarctica²⁴, and the $8\delta^{18}\text{O}_{\text{sw}}$ term corrects for the variation of δD with ice volume²⁹. The variation of $\delta^{18}\text{O}_{\text{sw}}$ as a function of Vostok age is taken from Sowers *et al.*¹⁶.

The new record shows a long cold period from ~ 140 to 200 kyr BP. During this cold period, the amplitude of temperature variation is smaller than during the last ice age. In central Antarctica this entire period was nearly as cold as the Last Glacial Maximum (LGM) ($\sim 6^\circ\text{C}$ colder than the Holocene). Temperature is warmer before ~ 200 kyr BP, with a well marked peak around 215 kyr BP.

Correlating Vostok and oceanic records

We now extend to the bottom of the core the previous efforts to correlate the Vostok and the SPECMAP timescales by curve-matching records of temperature^{17,30}, dust⁶, and $\delta^{18}\text{O}$ of O_2 and sea water^{12,16}. We plot changes in the isotopic composition of sea water, and hence inferred continental ice volume, against age (curve 3e).

The Vostok temperature record compares well with sea surface temperature (SST) in various Indian ocean cores^{17,30}. For example, in the summer SST record of RC11-120 (ref. 31), stages 7.1 and 7.3 are colder than the Holocene and much colder than

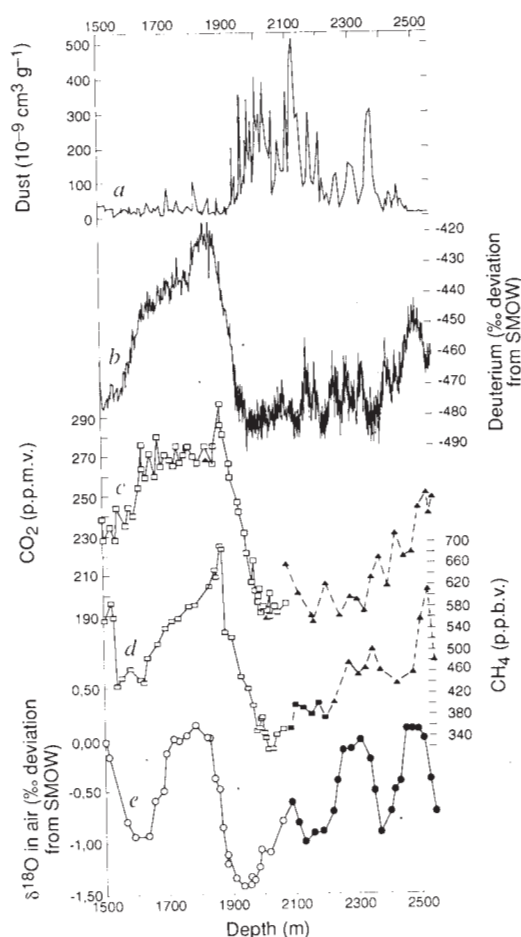


FIG. 1 Depth profiles a, Dust concentration. Data are from core 3 Γ down to 2,200 m (6) and from core 4 Γ below (discontinuous measurements every ~ 8 m). b, Deuterium content (δD in ‰ deviation from SMOW). In core 4 Γ , δD was measured continuously below 1,920 m on 1-m ice increments down to 2,200 m (~ 150 yr), at higher resolution (0.5 m, ~ 70 yr) down to 2,414 m, and on 1-m increments from this depth down to 2,546 m. Data from 3 Γ are reported for the 1,800–2,083-m depth interval. The agreement between 3 Γ and 4 Γ profiles is excellent over their common part. The comparison shows that there is no sizeable age difference between similar depth horizons in the cores despite small differences in the inclination of the two holes. c, d, e, Concentrations of CO_2 , CH_4 and $\delta^{18}\text{O}$ of O_2 within air bubbles. Individual measurements are indicated by symbols: open symbols correspond to published data, and filled symbols to new data; for CH_4 , squares and triangles correspond to different sets of measurements. Filled symbols are joined by a dotted line because the bottom part of the profile will undoubtedly evolve as additional samples are analysed.

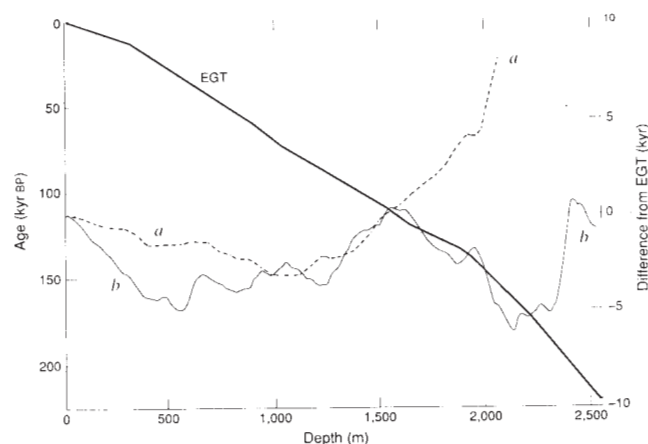


FIG. 2 The depth-age relationship for the EGT, together with differences between this and (a) the original timescale of Lorius *et al.*¹, and (b) the ^{10}Be timescale derived assuming a constant ^{10}Be flux and assigning the 110-kyr BP level at 1,534 m. Differences (right scale) are negative when the EGT gives older ages.

TABLE 1 Carbon dioxide, methane and oxygen

CO ₂ (p.p.m.v.)			CH ₄ (p.p.b.v.)			$\delta^{18}\text{O}$ (‰)		
Depth (m)	Age (kyr)	Mean	Depth (m)	Age (kyr)	Mean	Depth (m)	Age (kyr)	Mean
2078	150.1	197.7	2089	151.5	356	2089	151.6	0.60
2116	155.4	192.6	2102	153.6	400	2101	153.4	0.74
2157	160.3	189.7	2127	156.7	395	2109	154.5	0.82
2164	161.4	186.2	2153	159.8	382	2129	157.0	0.98
2203	166.9	205.1	2177	163.5	403	2139	158.1	0.89
2225	169.6	192.3	2196	166.1	376	2158	160.5	0.88
2247	172.9	184.5	2225	169.6	405	2167	161.9	0.80
2280	177.8	198.6	2273	176.9	478	2188	165.0	0.87
2302	180.9	198.5	2302	180.9	456	2198	166.3	0.79
2325	184.1	191.0	2325	184.1	468	2220	169.0	0.66
2333	185.3	190.8	2348	187.5	503	2233	170.7	0.38
2348	187.5	208.5	2373	191.1	464	2251	173.6	0.09
2363	189.6	214.0	2425	199.8	440	2278	177.6	0.11
2372	190.9	218.5	2475	207.5	460	2290	179.2	-0.13
2386	193.1	200.5	2501	211.5	558	2298	180.4	-0.10
2399	195.2	204.2	2525	214.9	612	2309	181.9	0.04
2414	197.9	211.5	2543	217.5	482	2329	184.7	-0.01
2425	199.8	232.2				2338	186.1	0.24
2437	201.6	232.3				2350	187.9	0.47
2451	203.7	218.8				2369	190.5	0.88
2475	207.5	220.9				2388	193.4	0.87
2499	211.2	243.5				2402	195.7	0.72
2525	214.9	251.9				2413	197.6	0.53
2533	216.0	240.0				2423	199.4	0.47
2543	217.5	248.6				2433	201.0	0.38
						2441	202.2	0.12
						2452	203.9	-0.11
						2462	205.5	-0.03
						2469	206.6	-0.11
						2489	209.7	-0.14
						2500	211.3	-0.06
						2510	212.7	0.04
						2518	213.8	0.20
						2528	215.3	0.39
						2544	217.6	0.66

The deuterium and dust data are available from the authors on request.

the Last Interglacial (curve 3a), in agreement with the Vostok temperature record. Correlating the Vostok temperature peak at 215 kyr BP with the RC11-120 peak at 218 kyr BP suggests that the bottom of the Vostok core corresponds to marine stage 7.4. The comparison with RC11-120 thus supports our glaciological timescale.

The dust concentration is plotted against EGT age in curve 3g. High values are observed throughout the penultimate glacial period (140–200 kyr BP). There is a small maximum at about 160 kyr BP and a broad minimum between 170 and 190 kyr BP. Low values are recorded before 200 kyr BP, consistent with our interpretation that this depth interval lies in the stage 7 interglacial period. We compare dust concentration at Vostok and the mass accumulation rate (MAR) in core RC27-61 from the Indian Ocean which gives a good record of aeolian input³². Both (curve 3f) are highest during glacial stages (2, 4 and 6), and low during the interglacial stages (1, 5 and 7). The two records are almost always in phase to within ± 10 kyr; the exception is between 170–190 kyr, where the low values in Vostok have no parallel in the RC27-61 record. At around 200 kyr BP, dust rises at Vostok 10 kyr before MAR rises at RC27-61, and at 145 kyr BP decreases at Vostok again about 10 kyr before the MAR decreases at RC27-61. It is not clear whether these differences reflect errors in the chronologies or true differences in phasing.

The $\delta^{18}\text{O}$ of past atmospheric O_2 is plotted against EGT age in curve 3d, after correction for differences between gas age and ice age³³, and for gravitational fractionation³⁴. Variations of $\delta^{18}\text{O}$ of O_2 are governed mainly by processes associated with oxygen isotope exchange of O_2 and sea water due to photosynthesis and respiration³⁵. These variations are thought to be closely

related to variations in the $\delta^{18}\text{O}$ of sea water: when the composition of sea water changes, the composition of O_2 produced by photosynthesis changes, transmitting the sea water variations to the atmospheric O_2 reservoir³⁶. For the sake of simplicity, we do not consider the small additional variations in the $\delta^{18}\text{O}$ of O_2 associated with glacial–interglacial variations in oxygen isotope fractionation by biogeochemical and hydrologic processes¹⁶ (that is, changes in the Dole effect).

The record of variations in $\delta^{18}\text{O}$ of O_2 looks similar to the record of variations in $\delta^{18}\text{O}$ of sea water (curve 3e). It also resembles the curve of June insolation at 20° N (curve 3c), because variations in $\delta^{18}\text{O}$ of sea water and variations in the Dole effect are both linked to insolation. We therefore test the EGT chronology by comparing $\delta^{18}\text{O}$ of O_2 with $\delta^{18}\text{O}$ of sea water and with insolation. Uncertainties restrict the comparison to about ± 10 kyr; within this uncertainty, variations in $\delta^{18}\text{O}$ of O_2 as dated by the EGT chronology are consistent with sea water $\delta^{18}\text{O}$ and insolation curve. Thus the $\delta^{18}\text{O}$ measurements, as well

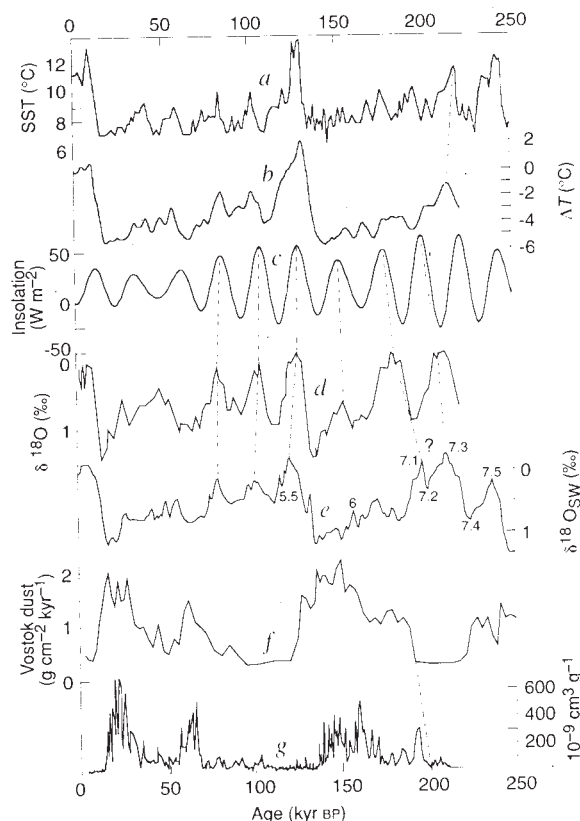


FIG. 3 Variations with time. a, Summer sea surface temperature (SST) at Indian Ocean site RC11-120 (16° 37' N, 59° 52' E). b, Atmospheric temperature change at Vostok (above the inversion) as derived from the deuterium profile. We used published deuterium data down to 1,920 m (ref. 2) and extended this record down to 2,083 m by averaging 31° and 41° data and to 2,546 m using 41° data alone; we then smoothed the entire record. c, Summer insolation at 20° N (ref. 49). d, $\delta^{18}\text{O}$ of O_2 measured in the air bubbles on the EGT. e, Change in the isotopic composition of sea water derived from the V19-30 $\delta^{18}\text{O}$ foraminifera record and extended back to 250 kyr BP¹⁶; the record is normalized so that the average Holocene value is 0‰. f, Mass accumulation rate (MAR) in the Indian Ocean site RC27-61 (adapted from ref. 32). g, The dust concentration in the Vostok ice. The Vostok records are reported using the EGT and the deep-sea records using the SPECMAP³¹ timescale. Horizons in the Vostok and deep-sea cores which are believed to be correlative are joined by dotted lines (see text).

as dust and temperature comparisons, support the EGT chronology.

The duration of the Last Interglacial, now highly controversial, is crucial for new insights on the causes of glacial–interglacial cycles^{37–39}. Relevant information is summarized in Fig. 4. Here we plot the Vostok temperature record using the adopted EGT and other timescales which would be derived from different starting assumptions. In the figure, we also show the $\delta^{18}\text{O}$ of sea water, the curve of sea surface temperature derived from core MD 84-551 (ref. 17) and the recently published³⁷ isotopic record of vein calcite from Devils Hole (Nevada). The LIG lasts twice as long at Vostok in the chronology of Lorius *et al.* than in the deep-sea record of the sea-level change^{1,2}.

Both the use of air $\delta^{18}\text{O}$ (ref. 16) and the correlation with Southern Ocean SST¹⁷ have recently confirmed that the temperature at Vostok rose before (~ 5 kyr at mid-transition) sea level rose for this termination (termination 2). As the EGT is, at mid-transition, ~ 4 kyr younger than that of Lorius *et al.*, the EGT and SPECMAP timescales now appear in agreement for this termination 2 (but differences of up to 6 kyr still exist at the LIG peak). The EGT timescale supports the observation, now well documented, that the duration of this interglacial is longer in the Vostok temperature and Southern Ocean SST records than in the record of sea-level change. This conclusion is fairly insensitive to our assumption about Vostok age at 1,534 m depth (a 5% change in age at 110 kyr BP will modify the duration by only 1 kyr or so). The duration which would be derived from the ^{10}Be chronology is similar, within 2 kyr, to that obtained with the EGT (but note the uncertainties of the ^{10}Be method mentioned above). Finally, the Vostok and Devils Hole records³⁷ agree less convincingly with the EGT than with the timescale of Lorius *et al.*¹.

Climate Interpretation

We now examine how the extension of the records, and the new information they contain, complement or modify our current climatic interpretation of Vostok data. In turn, we discuss the dust record, the link between orbital forcing and climate, and the record of greenhouse-gas concentrations.

The dust throughout the Vostok record begins to increase rapidly when Vostok atmospheric temperature drops by more than $\sim 4^\circ\text{C}$ below its modern value. We interpret this result as indicating that the penultimate glacial period, like the last glacial period, was characterized by more extensive deserts, more intense surface winds in the desert source regions and/or more efficient meridional transport. The isotopic composition from the LGM dust in the Dome C core⁴⁰ suggests that the Patagonia desert might be the main contributor to dust fallout in East Antarctica. During the end of the previous interglacial, as during the LIG and today, dust fallout was very low.

An intriguing feature of the records is that dust at Vostok covaries so closely with Vostok isotopic temperature whereas MAR at RC27-61 covaries so closely with sea water $\delta^{18}\text{O}$. The dust record is similar to the record of mass accumulation rate for the Northern Hemisphere Indian Ocean core RC27-61. Over the entire Vostok record, there are similarities between dust and indicators of dust source strength in Central Asia analysed in Chinese loesses^{41,42}. These similarities indicate that large-scale changes in dust fallout reflect global changes in climate.

We examine the link between Vostok temperature and orbital forcing through spectral analysis using a multi-taper method^{7,43}. The analysis was done (1) for the entire 2,546-m record, and (2) for the same series limited to 2,083 m using either the EGT or the Lorius *et al.* timescale. Using the EGT instead of the chronology of Lorius *et al.* causes a slight shift in the period of the obliquity peak (from 41.7 to 40.3 kyr), retains the strong significance of this peak and makes the significance of the precessional peak stronger, probably because of the shortening of the LIG. Using the full time series (EGT chronology) leads to a decrease in the amplitude of the obliquity peak (41.01 kyr) which is even

more significant; but the precessional peak (24.6 kyr) is now weaker and less significant. This decrease in amplitude of orbital frequencies is the quantitative expression of uniformly cold temperatures throughout the period between 140–200 kyr BP. During most of the record there is a strong relationship between local annual insolation, which primarily reflects obliquity, and temperature². This relationship is absent, however, between 140 and 200 kyr BP: the insolation maximum at ~ 170 kyr BP has no clear temperature counterpart.

Polar ice cores provide the most direct evidence of past changes in greenhouse gas concentrations. These changes are well documented for CO_2 and CH_4 and to a lesser degree for N_2O (ref. 44 for a review). Previously published results obtained on Vostok and other ice cores highlight the correlation between CO_2 concentration, CH_4 concentration and temperature throughout the last climate cycle. Results from Vostok also reveal phase differences between CO_2 and temperature. At the end of the penultimate interglacial, for example, the CO_2 decrease lags behind the Antarctic cooling leading to glacial conditions.

The CO_2 and CH_4 concentrations for the entire Vostok core profiles are plotted against age in Fig. 5 using the EGT after correction for the difference between gas age and ice age^{33,45}. CO_2 concentrations are low during the penultimate glacial period, from 140 to 190 kyr BP. These values are similar to the

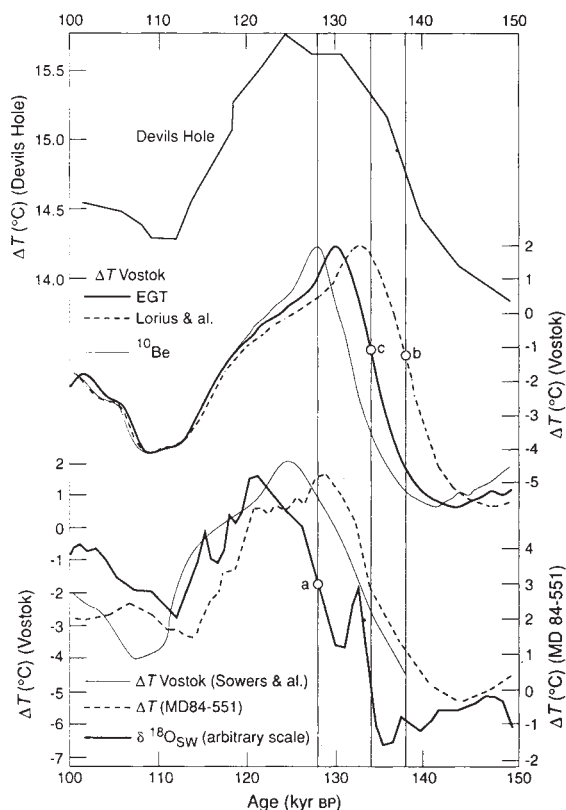


FIG. 4 The Vostok temperature record on various timescales for the 100–150 kyr BP period encompassing the LIG. Upper curve: the Devil's Hole record; middle curves: Vostok temperature using three different timescales (^{10}Be , EGT and Lorius *et al.*); lower curves: Vostok temperature (using the timescale obtained by Sowers *et al.*¹⁶ in correlating sea water $\delta^{18}\text{O}$ with ^{18}O of O_2); the sea surface temperature estimated from core MD84-551¹⁷; and the sea water $\delta^{18}\text{O}$ (similar to 3e with an arbitrary scale chosen in such a way this curve is comparable to the temperature records). The three vertical lines correspond to: a, mid-transition in sea water $\delta^{18}\text{O}$, a proxy of sea-level and ice volume change (128 kyr BP); b, mid-transition in Vostok temperature using the Lorius *et al.* timescale (138 kyr BP); and c, mid-transition according to the EGT reference timescale (134 kyr BP).

concentrations observed during the last glacial period between ~60 and 18 kyr BP. Before 190 kyr BP, the CO₂ values are higher. At the very bottom of the core, the CO₂ concentration is about 250 p.p.m.v., intermediate between full glacial concentrations and preindustrial values. The lowest CH₄ concentrations are similar to, or slightly lower than, those in the LGM, and are observed only during the final stage of the penultimate glaciation. The measurements are too scarce to discuss in detail this new part of the CH₄ record. Nevertheless, the increasing trend towards the bottom should be noted, especially the marked peak at the bottom with a maximum concentration of 610 p.p.b.v. Again, the concentrations measured near the very bottom are intermediate between full glacial and preindustrial values. This is consistent with the dating, indicating that the bottom ice is from the end of the previous interglacial period.

The strong covariation between CO₂ and temperature and CH₄ and temperature observed for the last 160 kyr is also present in the new part of the Vostok record. The CO₂-temperature correlation coefficient (r^2) over the past 220 kyr is 0.81 (0.76 for CH₄) instead of 0.78, in each case, over the last 160 kyr. In contrast to the lag of CO₂ behind temperature observed around 110 and 70 kyr BP, the CO₂ decrease between 220 and 190 kyr BP is in phase with the temperature change. A notable feature of the CH₄ results is the absence of clear oscillations during the penultimate glacial period. These would be expected from the link between the methane record and a monsoonal precipitation index largely dominated by the precession cycle⁴⁶ which was suggested for the past 160 kyr (ref. 9). This feature is directly reflected in the CH₄ spectrum: the amplitude and significance of the precession peak are smaller for the full record than for the last 160 kyr. The CO₂ spectrum also differs between the longer and shorter records, although in a different way. The 220-kyr

CO₂ record now shows significant peaks at obliquity and both precession periodicities, whereas only the precession peak was present for the last climate cycle⁷. This also reflects, in the spectral domain, what is clearly seen in the CO₂ record: the long penultimate glacial period and the last ice age have different characteristics. Obviously, all these features would need to be confirmed by further measurements.

Conclusions

Our new chronology appears to correlate well with oceanic records. In particular, the EGT chronology resolves half of the 10-kyr discrepancy in the duration of the LIG that arose when the previous Vostok chronology was compared to the deep-sea record. With the EGT chronology, the duration of the LIG at Vostok is still 5 kyr longer than its duration in deep-sea sediments. We believe that this important difference reflects the fact that, during termination 2, Antarctic temperatures rose ~5 kyr before the Northern Hemisphere ice sheets began to melt^{12,17}. In general, matching Vostok climate records with deep-sea sediments records and insolation curves supports our entire chronology to ± 10 kyr.

The isotopic temperature record shows that there is a long cold period in East Antarctica which we date from about 140 to 200 kyr BP. The coldest isotopic temperatures measured during this period are similar to those measured during the last glacial maximum, but the duration of the cold period is much longer. Isotopic temperatures rise in the bottom of the Vostok 4F core corresponding to interstadial-interglacial periods of marine stage 7. The overall covariation of climate variables in the newly examined section of the Vostok ice core is similar to that observed for the last glacial-interglacial cycle^{8,9}. Thus concentrations of CO₂ and CH₄ are lowest during the penultimate glacial period, where their values are comparable to those of the LGM. CO₂ and CH₄ are positively correlated with isotopic temperatures with correlations similar to those found during the last climate cycle. Throughout the record, dust concentrations at Vostok rise during cold periods, and are very low when temperatures are less than ~4 °C colder than modern values.

There is still much more climate information to extract from Vostok drilling. There is more than 1 km of ice below the 2,546 m reached by core 4F (the ice thickness is ~3,700 m), which may represent more than 500 kyr of time. At the same time, the new data discussed here, together with the results of the GRIP⁴⁷ and GISP2⁴⁸ cores, will allow comparison of climate change in Antarctica and Greenland beyond 200 kyr BP. □

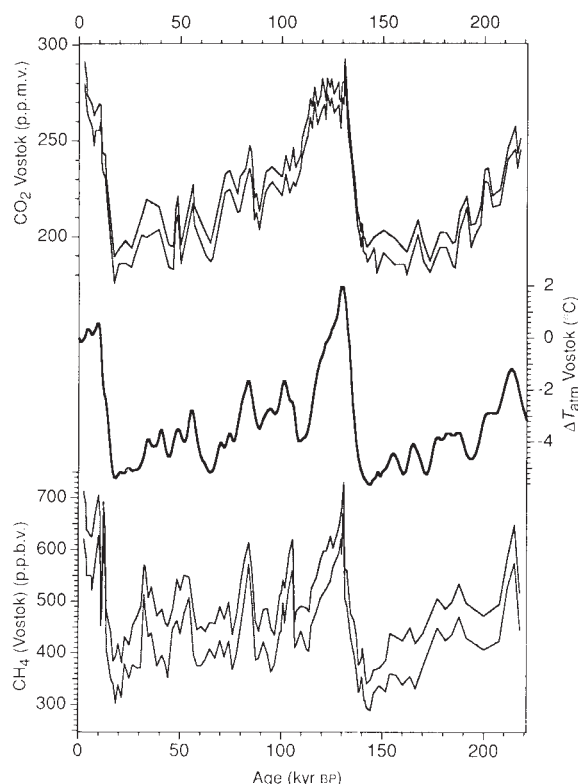


FIG. 5 CO₂, CH₄ and Vostok atmospheric temperature with respect to time (EGT) with the air-ice age difference calculated following Barnola *et al.*³³, taking into account the temperature dependence of the ice density⁵⁰. For CO₂ and CH₄, the envelope corresponds to the measurement accuracy.

Received 22 March; accepted 10 June 1993.

- Lorius, C. *et al.* *Nature* **316**, 591–596 (1985).
- Jouzel, J. *et al.* *Nature* **329**, 403–408 (1987).
- Yiou, F., Raisbeck, G. M., Lorius, C. & Barkov, N. I. *Nature* **316**, 616–617 (1985).
- Jouzel, J. *et al.* *Quat. Res.* **31**, 135–150 (1989).
- De Angelis, M., Barkov, N. I. & Petrov, V. N. *Nature* **325**, 318–321 (1985).
- Petit, J. R. *et al.* *Nature* **343**, 56–58 (1990).
- Yiou, P. *et al.* *J. geophys. Res.* **96**, 20365–20378 (1991).
- Barnola, J. M., Raynaud, D., Korotkevich, Y. S. & Lorius, C. *Nature* **329**, 408–414 (1987).
- Chappellaz, J., Barnola, J. M., Raynaud, D., Korotkevich, Y. S. & Lorius, C. *Nature* **345**, 127–131 (1990).
- Genthon, C. *et al.* *Nature* **329**, 414–418 (1987).
- Lorius, C., Jouzel, J., Raynaud, D., Hansen, J. E. & Le Treut, H. *Nature* **347**, 139–145 (1990).
- Sowers, T., Bender, M., Raynaud, D., Korotkevich, Y. S. & Orchado, J. *Paleoceanography* **6**, 679–696 (1991).
- Raisbeck, G. M. *et al.* *Nature* **326**, 273–277 (1987).
- Robin, G. de Q. *Phil. Trans. R. Soc. B* **280**, 143–148 (1977).
- Vaikmaa, R., Jouzel, J. & Petit, J. R., presentation at mtg on Application of Isotope Techniques in Studying Past and Current Environmental Changes in the Hydrosphere and the Atmosphere, Vienna, April 1993.
- Sowers, T. *et al.* *Paleoceanography* (submitted).
- Pichon, J. J. *et al.* *Paleoceanography* **7**, 289–318 (1992).
- Ritz, C. J. *Glaciol.* (in the press).
- Legrand, M. & Delmas, R. *Nature* **327**, 671–676 (1987).
- Raisbeck, G. M. *et al.* *The Last Deglaciation: Absolute and Radiocarbon Chronologies* (eds Bard, E. & Broecker W.S.) 127–139 (NATO ASI Series I2, Springer, Berlin, 1992).
- Raisbeck, G. M. & Yiou, F. *Nuclear Instrum. Meth.* **B5**, 91–99 (1984).
- Dansgaard, W. *Tellus* **16**, 436–468 (1964).
- Jouzel, J. & Merlivat, L. *J. geophys. Res.* **89**, 11749–11758 (1984).
- Lorius, C. & Merlivat, L. *IAHS Publ.* **118**, 127–137 (1979).
- Dahe, Q., Petit, J. R., Jouzel, J. & Stievenard, M. *J. Glaciol.* (in the press).
- Joussau, S. & Jouzel, J. *J. geophys. Res.* **98**, 2807–2830 (1993).

27. Jouzel, J., Koster, R. D., Suozzo, R. & Russell, G. EOS (Spring Meeting supplement, 7 April) (1992).
28. Johnsen, S. J. et al. *Nature* **359**, 311–313 (1992).
29. Craig, H. *Science* **133**, 1702–1703 (1961).
30. Shackleton, N. J., Mix, A. & Hall, M. A. *Quat. Sci. Rev.* **11**, 387–400 (1992).
31. Martinson, D. G. et al. *Quat. Res.* **27**, 1–29 (1987).
32. Clemens, S. & Prell, W. *Paleoceanography* **5**, 109–145 (1992).
33. Barnola, J. M., Pimentia, P., Raynaud, D. & Korotkevich, Y. S. *Tellus* **B43**, 83–90 (1991).
34. Craig, H., Horibe, Y. & Sowers, T. A. *Science* **242**, 1675–1678 (1988).
35. Dole, M., Lane, G. A., Rudd, D. P. & Zaukelies, D. A. *Geochim. cosmochim. Acta* **6**, 65–78 (1954).
36. Bender, M., Labeyrie, L. D., Raynaud, D. & Lorius, C. *Nature* **318**, 349–352 (1985).
37. Winograd, I. J. et al. *Science* **258**, 255–260 (1992).
38. Lambeck, K. & Nakada, M. *Nature* **357**, 125–128 (1992).
39. Imbrie, J., Mix, A. C. & Martinson, D. G. *Nature* **363**, 531–533 (1993).
40. Grousset, F. et al. *Earth planet. Sci. Lett.* **111**, 175–182 (1992).
41. Kukla, G., An, Z. S., Melice, J. L., Gavin, J. & Xiao, J. L. *Trans. R. Soc. Edin.* **81**, 263–288 (1990).
42. Beer, J. et al. *Geophys. Res. Lett.* **20**, 57–60 (1993).
43. Thomson, D. J. *Proc. IEEE* **70**, 1055–1096 (1982).
44. Raynaud, D. et al. *Science* **259**, 926–934 (1993).
45. Sowers, T. A., Bender, M. L., Raynaud, D. & Korotkevich, Y. L. *J. geophys. Res.* **97**, 15683–15697 (1992).
46. Prell, W. & Kutzbach, J. E. *J. geophys. Res.* **92**, 8411–8425 (1987).
47. Dansgaard, W. et al. *Nature* **364**, 218–220 (1993).
48. Taylor, K. T. et al. *Nature* **361**, 432–436 (1993).
49. Berger, A. L. *J. atmos. Sci.* **35**, 2362–2367 (1978).
50. Martinerie, P., Raynaud, D., Etheridge, D., Barnola, J. M. & Mazaudier, D. *Earth planet. Sci. Lett.* **112**, 1–13 (1992).

ACKNOWLEDGEMENTS. Vostok is a joint project between Russia, France and the United States. We thank participants in drilling, field work and ice sampling. We acknowledge the Russian Antarctic Expeditions, the Institut Français de Recherches et Technologies Polaires and the NSF Division of Polar Programs for logistic support. The project is supported in Russia by the Russian Ministry of Sciences, in France by the Programme National d'Études de la Dynamique du Climat, Fondation de France and the Commission of European Communities Environment Programme, and in the U.S., by the NSF. We thank J. Meyssonier for discussions, J. C. Pugno for the figures and C. Alba, P. Doira and M. Delmotte for isotope analyses.

Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5

Kirk L. Clark^{*†}, Elaine D. Halay^{*†}, Eseng Lai[†] & Stephen K. Burley^{*†§}

^{*} Laboratories of Molecular Biophysics and [†] Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

[‡] Cell Biology and Genetics Program, and Endocrinology Service Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA

The three-dimensional structure of an HNF-3/fork head DNA-recognition motif complexed with DNA has been determined by X-ray crystallography at 2.5 Å resolution. This α/β protein binds B-DNA as a monomer, through interactions with the DNA backbone and through both direct and water-mediated major and minor groove base contacts, inducing a 13° bend. The transcription factor fold is very similar to the structure of histone H5. In its amino-terminal half, three α -helices adopt a compact structure that presents the third helix to the major groove. The remainder of the protein includes a twisted, antiparallel β -structure and random coil that interacts with the minor groove.

MEMBERS of the HNF-3/fork head family of eukaryotic transcription factors share a highly conserved DNA-binding domain and occur in various organisms ranging from yeast to man (reviewed in ref. 1). Hepatocyte nuclear factor-3 (HNF-3) proteins were first identified as activators of liver-specific gene expression in rat². Initial evidence for a new DNA-binding motif came when HNF-3 α , β and γ were found to share a novel, highly conserved DNA-binding region^{3,4} with the nuclear protein product of the *Drosophila* homeotic fork head gene, involved in the formation of terminal structures of the early fly embryo^{5,6}.

Functionally significant HNF-3 binding sites were identified in the promoters of a large number of genes expressed in mammalian hepatocytes (for example, albumin⁷, phosphoenolpyruvate carboxykinase⁸ and two liver-specific transcription factors, HNF-1 α (ref. 9) and HNF-3 β (ref. 10)), indicating roles for HNF-3 in both cell differentiation and tissue-specific gene expression. Other members of the HNF-3/fork head family have also been implicated in development; fork head mutant embryos lack structures contributing cells to the fore- and hindgut of the fly⁵. The sloppy paired proteins (slp1 and slp2) affect fly segmentation¹¹. XFKH1 is expressed in the blastopore-lip of *Xenopus* embryos and is essential for normal axis formation¹².

We now report the X-ray crystallographic structure determin-

ation of a canonical HNF-3/fork head DNA-recognition motif bound to its target DNA sequence. The three-dimensional structure of the DNA-binding domain of HNF-3 γ bound to its cognate DNA, obtained at 2.5 Å resolution, reveals a monomeric α/β protein structure interacting with B-DNA. The N-terminal half of the polypeptide chain adopts a compact three-helix structure, which presents its third α -helix to the major groove. The carboxy-terminal half of the protein consists of β -strands and random coil and makes various DNA contacts. Elucidation of this DNA-recognition fold provides a three-dimensional scaffold with which to understand the DNA-binding properties of members of the HNF-3/fork head family of transcription factors. Structural comparisons demonstrate a striking similarity to an unrelated, DNA-packaging protein histone H5 (ref. 13).

Structure determination and refinement

A fragment of HNF-3 γ , HNF-3 γ (107–223), containing the conserved DNA-binding region, was expressed, purified and co-crystallized with a 13-base-pair (bp) duplex oligonucleotide bearing a sequence derived from its transthyretin (TTR) promoter binding site² (Fig. 1 and Table 1). Quantitative electrophoretic mobility shift (EMS) assays² documented specific nanomolar-affinity DNA binding of the purified recombinant protein with the sequence used for co-crystallization at a protein:

§ To whom correspondence should be addressed.

DNA duplex stoichiometry of 1:1 (data not shown). The trigonal cocrystals (space group $P3_1$; $a=74.6$ Å, $c=85.05$ Å) diffract anisotropically to 2.2 Å along the cylindrical axis of the DNA (nearly coincident with c^*) and to about 2.6 Å in the other two axial directions. The asymmetric unit includes two DNA-protein complexes.

The structure was solved at 2.84 Å resolution by multiple isomorphous replacement (MIR) using four singly labelled, iodinated-DNA derivatives (Table 1). There was considerable non-isomorphism between derivative and true native co-crystals, requiring phase determination with 'native' diffraction data measured from a co-crystal of HNF-3 γ (107-223) complexed with singly brominated DNA (Table 1). The heavy-atom sites were located by difference Patterson analyses, and refinement against both isomorphous and anomalous differences gave a final figure of merit of 0.55 (ref. 14). The MIR map revealed strong electron density for the phosphodiester backbone, most of the bases, three right-handed α -helices and three β -strands in each half of the asymmetric unit. 'Solvent flattening' with SQUASH¹⁵ permitted two standard B-DNA models (one for each half of the asymmetric unit), with the sequence shown in Fig. 1*b*, to be positioned in the electron density map by overlaying thymine methyl carbon atoms on the corresponding iodine atomic coordinates, with root-mean-square (r.m.s.) deviation of

1.0 Å. Thereafter, the DNA was manually rebuilt¹⁶ to fit the modified electron density, and the DNA model was further improved using XPLOR positional refinement¹⁷. Phase combination with the MIR data and the refined DNA model¹⁸ revealed continuous electron density for the N-terminal half of the polypeptide chain, which was initially built as three α -helical polypeptide segments. Several rounds of model building and phase combination interspersed with positional refinement allowed an unambiguous trace and sequence assignment of the polypeptide chain.

Estimation of and correction for the overall anisotropic temperature factor allowed refinement of the atomic model at successively higher scattering angles. At 2.5 Å resolution, the model was improved using simulated annealing¹⁹ and positional refinement with dynamically constrained deoxyribose puckers (C_2 -endo). Omit $(2|F_{\text{observed}}| - |F_{\text{calculated}}|)$ difference Fourier syntheses were calculated with the refined model and several rounds of manual rebuilding and positional refinement were done to improve stereochemistry. Finally, tightly restrained, individual isotropic B-factors were refined and well ordered solvent molecules were included. The current model consists of two copies of the protein-DNA complex (13-bp oligonucleotide depicted in Fig. 1*b* and the HNF-3 γ (107-223) polypeptide sequence from 113 to 218), two Mg^{2+} ions and 131 solvent mol-

a

	120	130	140	150	160	170	180	190	200	210	
HNF-3 γ (117)	HAKPPYSISLITMAIQAPGKMLTSEIYQWIMDLFPYYRENQQRWQNSIRHLSFNDCFKVARS	DPDKP	-GKGSY	WALHPSSGNMF	ENG	YCY	--LRR	QKRFKL			
HNF-3 α (168)		S		Q		A		-	T	D	C 95%
HNF-3 β (157)		S N		F Q			L P A	-	F T	D	C 90%
XFKH1 (99)		N M N	V	Q			I P E	-	T E		C 90%
fkh (208)		NN TR	F	F Q			IP T	-	F T	D	D 88%
BF-1 (170)	YE F NA M	R S E R	NG EF KN	K G		N L K	P HY D	-	N M D	DDV IG TTGK	RSTTSR 55%
slp2 (178)	NE NA M	R SSE R	NG EY TNH	D K G		N L K	P HY D	-	N M D	AEDV IG STGK	RTTAAS 53%
slp1 (118)	TK NA M	DS EQR	NG YLINR	FKA KRG		N L K	T IP Y D	-	N I D	AEV IGETT	KNPGAS 48%
ILF (139)	DS AQ VQ	TM D Q	NG TH TKNY	TADKG		N L RY I	P QEE	-	F RID A	ESKLIEQAF--RK RP GVP	47%
HTLF (20)	TS FSL Y	EHS N C PVK	S L H	FATAPTG K	V N	L K Q	E HG VN		L CVD	EYKPNLI-QA---KK PFSSA	45%
Yfkh (105)	AK AT CL	L SQEGK	Q H HVH	KQK DAS		N L A I	TEK C	-	HF EVR	GAETK FK EN--RGYEFVKDS	47%
Consensus	KPPYSY ALI MAI	A K L T L S I Y W I	FPYYR N	WQNSIRH	LS N C F	KVPRS	DKP-GKGSY	W L P S	F G	--LRR	
Environment	pepppbppbbppbbpp	peppbbppbbppbbpp	peppbbppbbppbbpp	peppbbppbbppbbpp	peppbbppbbppbbpp	peppbbppbbppbbpp	peppbbppbbppbbpp	peppbbppbbppbbpp	peppbbppbbppbbpp	peppbbppbbppbbpp	peppbbppbbppbbpp
DNA Binding Residues	P		R		P B B	PWW		P P		RBP W	
Secondary Structure	-----HHHHHHHHHHH	-----SSSSSSSSSSSS	-----HHHHHHHHHHH	-----HHHHHHHHHHH	-----SSSSSSSS	-----SSS	-----SSS	-----SSS	-----SSS	-----SSS	-----SSS
	H1	S1	H2	T'	H3	Mg	S2	W1	S3	W2	

b

	-111	-85
TRANSTHYRETIN PROMOTER	(5') GTTGACTAAGTCAATAATCAGAATCAG (3')	(3') CAACTGATTGATTATAGTCTTAGTC (5')
HNF-3 FOOTPRINT DATA	CC*****DDDD	
	1	13
CRYSTALLIZATION DNA	(5') GACTAAGTCAACC (3')	(3') CTGATTGATTGG (5')

FIG. 1 *a*, Partial sequence alignment of the HNF-3/*fork head* DNA-binding domains, including HNF-3 γ ⁴, HNF-3 α ⁴, HNF-3 β ⁴, XFKH1¹², fkh⁵, BF-1⁴⁰, slp2¹¹, slp1¹¹, ILF⁴¹, HTLF²⁶ and Yfkh⁴². Amino acids (in one-letter code) are numbered according to the sequence of HNF-3 γ (GenBank accession number L09648). The secondary structural elements of HNF-3 γ (107-223) are denoted using H for α -helix and S for β -strand and the labelling convention of Fig. 3*a*. The degree of sequence identity between each DNA-binding domain and that of HNF-3 γ is given as a percentage. Consensus amino acids and those involved in mediating phosphate (P), ribose (R), base (B) and water-mediated base (W) contacts with DNA are indicated. The environment of each residue in the isolated protein, as judged by Eisenberg's 3D profile analysis²², is given

as either exposed (e), partially buried (p) or largely buried (b). These categories are based on burial of total surface area not fraction of surface area, and the largely buried category (b) includes some bulky amino acids such as tyrosine and arginine that are mostly buried but interact with DNA through exposed, polar functional groups. *b*, Sequence of the duplex oligonucleotide used in cocrystallization and its relationship to the so-called strong transthyretin (TTR) promoter binding site². The crystallographic DNA numbering scheme is shown with reference to the numbering of the TTR promoter. The DNA footprinting data are illustrated schematically, with C and D denoting free-radical and DNase I protection, respectively. Asterisks represent sites at which HNF-3 γ protects DNA against both methods.

TABLE 1 Summary of crystallographic analysis

	'Native' BrdU2	Native	IdU8	IdU2	IdU6	IdU8
Resolution (Å)	2.5	2.8	2.8	3.8	2.65	2.7
Reflections (observed/unique)	100,640/18,352	46,199/13,139	38,735/13,139	16,000/5,237	55,480/15,485	37,668/14,502
Data coverage (%)	98.5	93.6	87.2	92.0	93.7	89.4
$R_{\text{sym}}(I)$ (%) ($I > 0$)	6.1	7.0	8.8	7.2	7.7	7.4
Mean fractional isomorphous difference (%)						
versus native			16.9	17.1	17.0	19.0
versus 'native' BrdU2			21.7	14.0	13.8	20.0
MIR analysis (15–2.84 Å)						
Phasing power versus resolution (Å)						
15.0–8.66			1.89	1.00	1.20	2.10
8.66–5.89			2.08	0.86	1.35	2.25
5.89–4.73			1.85	0.71	1.29	2.18
4.73–4.06			0.79	0.45	0.70	1.25
4.06–3.62			0.45		0.45	0.58
3.62–3.29			1.44		1.09	1.39
3.29–3.04			1.35		1.27	1.37
3.04–2.84			1.41		1.24	1.42
Overall phasing power			0.97	0.69	0.86	1.23
Mean overall figure of merit	0.55					
Refinement (6–2.5 Å)						
R factor (%)	21.1					
Reflections ($ F > 1\sigma F $)	18,352					
Total number of complex atoms	2,806					
Total number of magnesium ions	2					
Total number of solvent molecules	131					
r.m.s. bond length (Å)	0.017					
r.m.s. bond angle (degrees)	2.45					
r.m.s. B-factor bonded atoms (Å ²)	3.10					

The DNA-binding domain of rat hepatocyte nuclear factor-3 γ consisting of amino acids 107–223 plus three extra amino acids Lys-Ser-Ile, HNF-3 γ (107–223), was overexpressed in *Escherichia coli* (BL21(DE3)pLysE) using the T7 RNA polymerase system⁴⁵. The amino-acid numbering system used here differs from that reported in ref. 4, and corresponds to the definitive amino-acid sequence given in GenBank accession number L09648. Cells were grown at 30 °C to an absorbance of 0.6 at 595 nm and induced with 0.5 mM isopropyl- β -D-thiogalactopyranoside, followed 30 min later by addition of rifampicin. Three hours after induction, the cells were collected by low-speed centrifugation, and lysed by three cycles of freeze-thaw in the presence of 0.1% Nonidet P-40 and DNaseI. Purification and detergent removal were achieved by 3 cation-exchange chromatographies at 2 pHs, resulting in protein with an estimated purity of better than 98%. High-resolution mass spectrometry⁴⁶ showed that the purified, recombinant protein lacked the N-terminal methionine and the last two amino acids (Ser-Ile), but was otherwise intact. Further control experiments ensured that the protein used for crystallization was biochemically active. Electrophoretic mobility shift assays (as in ref. 2) documented full DNA-binding activity under reducing conditions. DNA was synthesized using standard phosphoramidite chemistry, and purified by anion exchange chromatography at pH 10.0 and 55 °C. Single-stranded DNA was quantified by UV spectrophotometry using extinction coefficients calculated incorporating nearest-neighbour contributions⁴⁷, mixed with an equimolar amount of complementary strand, and annealed by heating to 65 °C for 30 min and then cooling to 4 °C at a uniform rate of 0.33 °C per min. HNF-3 γ (107–223)-DNA complex was prepared by mixing concentrated binding domain with concentrated, annealed, double-stranded DNA to a final concentration of 1 mM in 100 mM KCl, 550 mM ammonium acetate, 2 mM MgCl₂, 20 mM dithiothreitol, 20 mM potassium acetate, pH 5.5. The DNA-protein complex was relatively insoluble under low-salt conditions, enabling crystals to be grown at 4 °C by vapour diffusion against a reservoir containing 50 mM KCl, 2 mM MgCl₂, 20 mM dithiothreitol, 100 mM potassium acetate pH 5.5. Hexagonal bars grew over the course of weeks to months to typical diameters of 0.2 mm and lengths of over 1 mm. Crystals of heavy-atom derivatives were prepared in the same manner using DNA in which 5-iodoU or 5-bromoU had been substituted for T in the positions denoted with the numbering scheme of Fig. 1b. Care was exercised during the anion exchange purification step to remove dehalogenated DNA using the pK_a difference between bases with and without the halogen. X-ray data collection at conventional sample temperatures (4 to 20 °C) was complicated by relatively weak diffraction and severe radiation sensitivity. Diffraction measurements were at –150 °C using a Rigaku RAXIS-IIc imaging plate area detector and a refrigerated nitrogen gas delivery system. Crystals were transiently 'cryoprotected' in a 30% glycerol, 10% PEG4K, 10 mM MgCl₂ mixture before placement in a 1-mm diameter wire loop and freezing. Significant increases in mosaic spread and corresponding losses of diffraction intensity were eliminated by optimization and standardization of the freezing protocol. Despite isolation of the crystal and area detector in a nitrogen filled chamber and frequent 'dusting' of the crystal, modest accumulations of frost on the frozen crystal and supporting loop produced an X-ray powder line of ice at about 3.8 Å resolution, reducing heavy-atom derivative phasing power in the resolution shell between 4.0 and 3.6 Å resolution to below 1.0. Oscillation photographs were reduced to structure factor amplitudes and scaled using DENZO and SCALEPACK (Z. Otwinowski, personal communication). The 'native' BrdU2 diffraction data used for structure refinement was 87% complete with $\langle I/\sigma(I) \rangle = 6.5$ and $R_{\text{sym}} = 16.6\%$ in the highest resolution shell between 2.58 and 2.5 Å. Additional crystallographic software included the PROTSYS package (G. A. Petsko, personal communication), COMBINE¹⁸, DSCALEAD (M. Rould, personal communication) and NEWHEL92 (R. E. Dickerson, personal communication). Atomic coordinates and structure factor amplitudes have been submitted to the Brookhaven Protein Data Bank⁴⁸ and will be held for 1 and 4 years, respectively. (Atomic coordinates are available immediately to academic users from S.K.B.; non-academic users should contact the Office of Technology Transfer at The Rockefeller University.)

$R_{\text{sym}} = \sum |I - \langle I \rangle| / \sum I$, where I = observed intensity, $\langle I \rangle$ = average intensity obtained from multiple observations of symmetry related reflections.

Mean fractional isomorphous difference = $\sum ||F_{\text{PH}}| - |F_{\text{P}}|| / \sum |F_{\text{P}}|$, $|F_{\text{P}}|$ = protein structure factor amplitude, $|F_{\text{PH}}|$ = heavy-atom derivative structure factor amplitude.

Phasing power = $\text{r.m.s.}(|F_{\text{H}}|/E)$, $|F_{\text{H}}|$ = heavy atom structure factor amplitude and E = residual lack of closure.

R.m.s. bond lengths and r.m.s. bond angles are the respective root-mean-square deviations from ideal values; r.m.s. thermal parameter is the root-mean-square deviation between the B values of covalently bonded atomic pairs.

ecules. Refinement statistics are given in Table 1. The electron density for the polypeptide backbone is continuous everywhere at 1.2σ in a $(2|F_{\text{observed}}| - |F_{\text{calculated}}|)$ difference Fourier synthesis, except for the N-terminal residues 113 to 116 and breaks in two loops at 182 to 188 and 203 to 206 in one complex (representative portions of the final electron density are shown in Fig. 2). The other complex in the asymmetric unit is somewhat less well ordered. A Ramachandran analysis²⁰ revealed 6 of 210 backbone torsionangle combinations in unfavourable regions of the (ϕ, ψ) plot (data not shown). The electron density for the DNA is well defined throughout the asymmetric unit.

Structure description

Topological overview. The three-dimensional structure of the HNF-3 γ (107–223)–DNA complex is illustrated in Fig. 3. The polypeptide chain binds DNA as a monomer, adopting a compact α/β structure with its N and C termini on opposite faces of the molecule. Three α -helices (H1, H2 and H3) dominate the N-terminal half of the protein. Between H1 and H2 is a β -strand (S1), which interacts with strands S2 and S3 to form a three-stranded, twisted, antiparallel β -sheet. The T' loop connecting H2 and H3 runs along the phosphodiester backbone. Strands S2 and S3 of the β -sheet are connected by a loop (W1), interacting with the phosphate backbone of a neighbouring DNA molecule (Fig. 3d). An extended polypeptide chain emerges from the C terminus of S3 to form another long loop (W2), making minor and major groove contacts and disappearing in a solvent channel.

Hydrophobic core. Although the N-terminal half of HNF-3 γ (107–223) is topologically similar to the three α -helices of the homeodomain proteins (reviewed in ref. 21), it is intimately associated with the C-terminal half of the protein and does not

represent a readily separable subdomain of the protein. The strand S1, connecting H1 and H2, is the first member of the antiparallel β -sheet. Moreover, the fold generates a well defined hydrophobic core that includes the entire polypeptide chain. The structure of the isolated protein was analysed using Eisenberg's 3D profile method²² and the environment of each residue is given in Fig. 1a.

Hydrophobic surface patch. The fold of the HNF-3 γ (107–223) polypeptide chain creates a large, superficial hydrophobic patch derived from exposure of Tyr 122 in H1, Trp 148 in H2, Phe 153 in T', Phe 203 and Tyr 208 in W2, which make van der Waals and weakly polar interactions²³ with the side chain of Met 130 in H1 to give a rosette-like structure (Fig. 2a). With the exception of Tyr 208, the aromatic residues surrounding Met 130 are either conserved aromatic or conserved hydrophobic residues (Fig. 1a). This hydrophobic surface feature is, therefore, almost certainly conserved within the HNF-3/*fork head* family and represents an unusual combination of sulphur–aromatic²⁴ and aromatic–aromatic interactions²⁵.

W1 and W2 loops. Loop W1 connects strands S2 and S3 and interacts with the DNA. Although the end of W1 is probably disordered in the absence of DNA, the loop stem is stabilized by β -sheet hydrogen bonding between S2 and S3 and side chain–backbone interactions, for example Arg 180 to Gly 190. Loop W2 emerges from the C terminus of S3 and meanders across the surface of the three-helix bundle in the vicinity of the N terminus of α -helix H1 and the C terminus of α -helix H3, completing the hydrophobic core of the protein by burying Leu 195, Met 202, Arg 211 and Phe 215 in W2. A 3' deletion in the gene for HNF-3 α that eliminated W2 beyond Ser 250 (or Ser 199 in HNF-3 γ) abolished DNA binding in EMS assays³.

Mg²⁺ binding site. Figure 4 illustrates the protein–DNA inter-

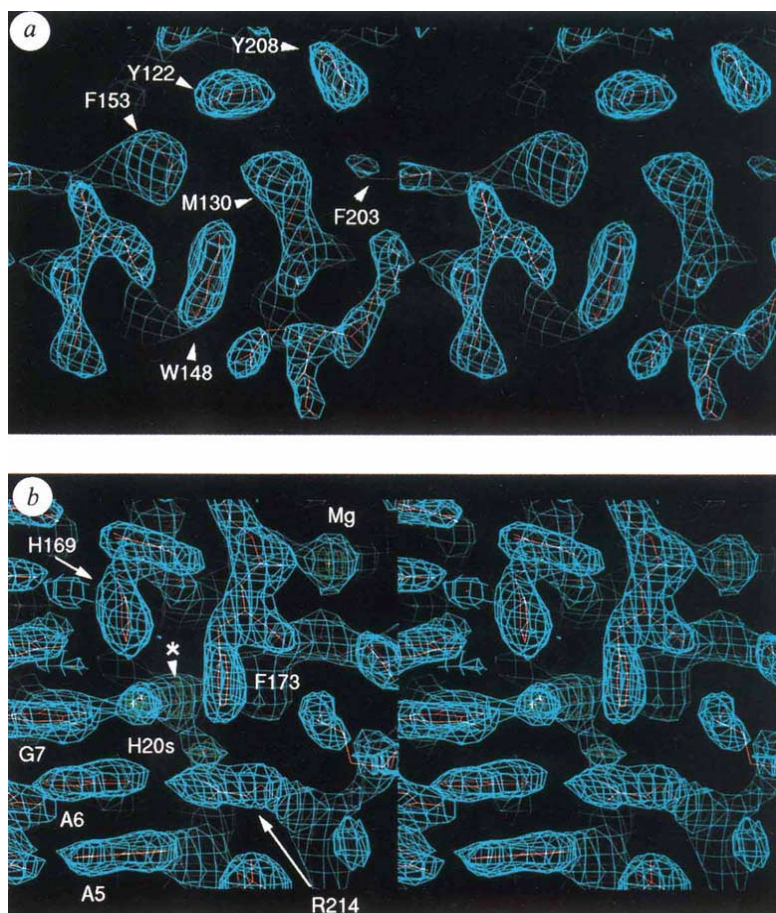
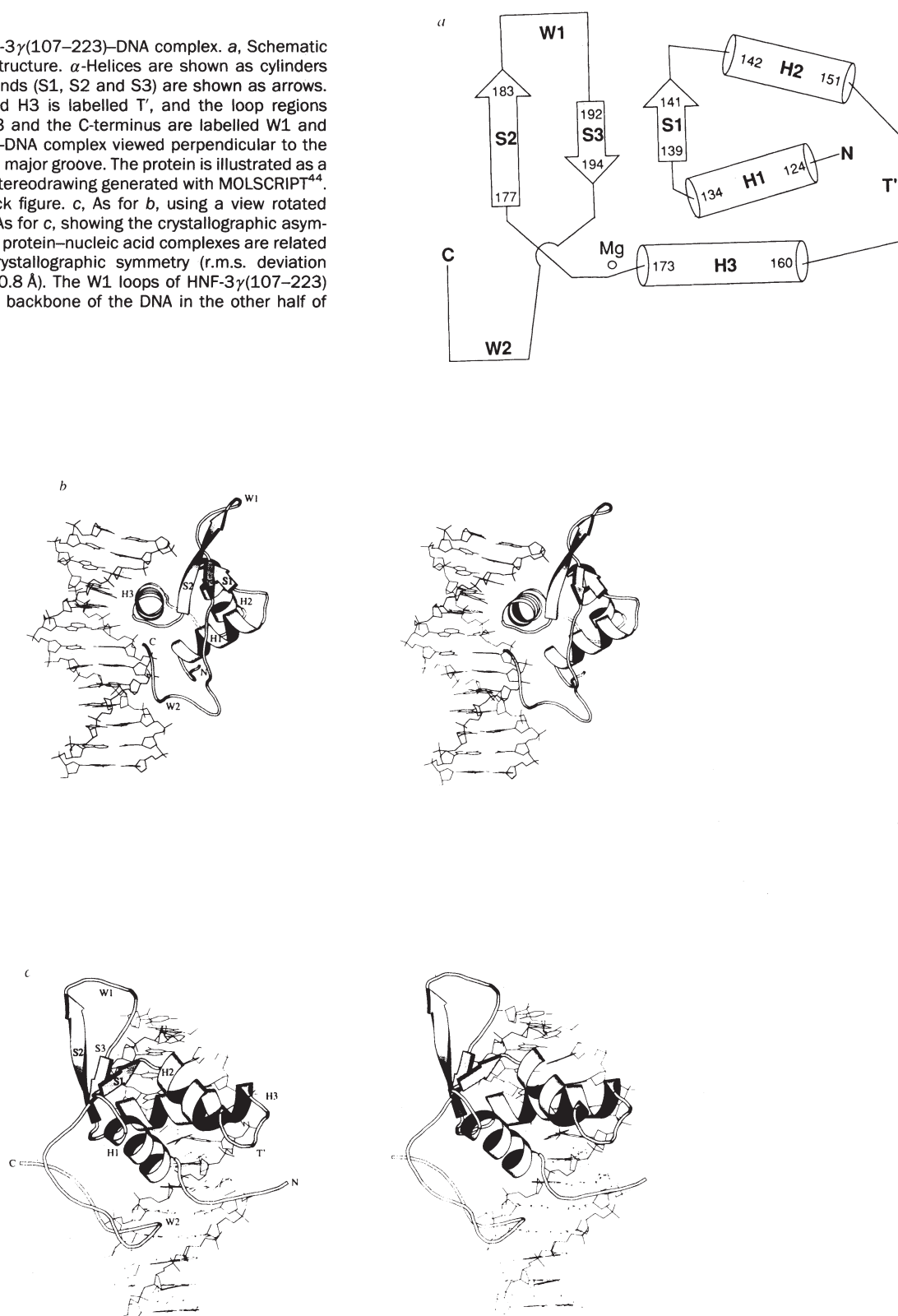


FIG. 2 Stereodrawings of representative portions of the final electron density map generated using the program O¹⁶. The results of a $(2|F_{\text{observed}}| - |F_{\text{calculated}}|)$ difference Fourier synthesis are shown in blue, contoured at 1.7σ , and superimposed with a stick representation of the atomic model in red. **a**, The hydrophobic surface patch formed by the aromatic-methionine rosette (see text for details). **b**, Helix H3 interacting with the major groove of DNA. This view includes $(|F_{\text{observed}}| - |F_{\text{calculated}}|)$ omit difference electron density for the water molecules (H_2O) bound in the major groove and the Mg^{2+} ion, coloured green and contoured at 5 and 10σ , respectively.

FIG. 3 Structure of the HNF-3 γ (107–223)–DNA complex. *a*, Schematic drawing of the secondary structure. α -Helices (H1, H2 and H3) and β -strands (S1, S2 and S3) are shown as arrows. The region between H2 and H3 is labelled T', and the loop regions between S2 and S3 and S3 and the C-terminus are labelled W1 and W2, respectively. *b*, Protein–DNA complex viewed perpendicular to the axis of α -helix H3 lying in the major groove. The protein is illustrated as a Richardson-type⁴³ cartoon stereodrawing generated with MOLSCRIPT⁴⁴. The DNA is drawn as a stick figure. *c*, As for *b*, using a view rotated $\sim 90^\circ$ about the vertical. *d*, As for *c*, showing the crystallographic asymmetric unit, in which the two protein–nucleic acid complexes are related by a 2-fold axis of non-crystallographic symmetry (r.m.s. deviation between α -carbon atoms = 0.8 Å). The W1 loops of HNF-3 γ (107–223) interact with the phosphate backbone of the DNA in the other half of the asymmetric unit.



action and the environment of a well ordered Mg^{2+} ion (refined temperature factor = $4 \text{ \AA}^2$). The divalent metal is coordinated by four carbonyl oxygen atoms from the polypeptide backbone, including $\text{O}(\text{Leu } 171)\text{-Mg}^{2+} = 2.5 \text{ \AA}$, $\text{O}(\text{Asn } 174)\text{-Mg}^{2+} = 2.5 \text{ \AA}$, $\text{O}(\text{Phe } 177)\text{-Mg}^{2+} = 2.5 \text{ \AA}$, and $\text{O}(\text{Ser } 172)\text{-Mg}^{2+} = 3.2 \text{ \AA}$, and there is electron density corresponding to two appropriately placed water molecules, giving six occupied coordination sites in all. This tightly bound Mg^{2+} ion appears to stabilize the compact structure of the protein by tethering the C terminus of α -helix H3 to the N terminus of strand S2.

DNA structure. The asymmetric unit is illustrated in Fig. 3d. Like many cocrystals of nucleic acid-protein complexes, the DNA in these crystals forms pseudocontinuous helices. Within the asymmetric unit, two colinear oligonucleotides are rotated and horizontally displaced with respect to one another giving a 3' to 3' arrangement, partially stabilized by C-C base stacking ($3.4 \text{ \AA}$). Analysis of the structures of the oligonucleotides themselves, reveals some deviations from canonical B-form DNA. The average helical twist is 35° and the mean rise per base pair is $3.2 \text{ \AA}$, implying 10.3 bp per turn. There is a bend of about 13° , narrowing the major groove in which H3 is located.

Comparison of HNF-3/fork head sequences

Twenty-five per cent of the 101 amino acids constituting the HNF-3/fork head DNA-recognition motif are absolutely conserved among all the family members in species examined thus far (Fig. 1a). Only two short insertions have been found and these map, respectively, to loop regions W1 and W2 of the three-dimensional structure of HNF-3 γ (107-223). Between HNF-3 γ and other family members (Fig. 1a) the degree of sequence identity ranges from 45% (HTLF²⁶) to 95% (HNF-3 α ³). The sites at which differences occur map either to the surface of the molecule, where mutations would be tolerated, or represent conservative changes of buried residues unlikely to destabilize the protein's hydrophobic core (Fig. 1a gives a summary of amino-acid environments). This level and pattern of sequence conservation indicate that all known members of the HNF-3/fork head family must share the same three-dimensional structure in the DNA-recognition region²⁷. Subsequent references to the results of

DNA-binding studies use the HNF-3 γ sequence numbers throughout (Fig. 1a gives conversions to other sequence numbering schemes).

DNA binding and promoter recognition

HNF-3 γ (107-223) interacts with DNA as a monomer, over a linear distance of about $40 \text{ \AA}$ along the axis of the double helix (Fig. 3d). This extensive protein-nucleic acid interaction surface compares favourably with the results of free-radical and DNase I footprinting studies of HNF-3 γ bound to the TTR promoter^{2,3} (Fig. 1b). Contacts between protein and DNA occur on both DNA strands and are distributed throughout the linear sequence of the polypeptide chain. α -helices H1 and H2 are oriented with their N termini pointing towards the negatively charged phosphate backbone of the DNA. The principal contact surface is provided by α -helix H3, which lies in the major groove and is flanked by two long loops (W1 and W2) that touch DNA backbone and base atoms. The structure of HNF-3 γ (107-223) can be likened to a butterfly with an α -helical thorax and two wing-like loops, and might usefully be referred to as the 'winged-helix' motif.

Details of the interactions between HNF-3 γ (107-223) and the DNA are depicted in Fig. 4. Contacts with the DNA backbone involve both phosphate and ribose groups, and are typical of protein-DNA complexes²¹. On the sense strand of the TTR promoter these interactions are restricted to base positions 5 to 9, and on the antisense strand they include 1 and 2, and 11 to 13 (Fig. 4c legend gives details).

Side chain-base contacts are restricted to the sense strand of the duplex oligonucleotide and extend between positions 4 and 10 (see Figs 1b, 2b and 4a, b, c). In the minor groove, Arg 210 makes a bidentate interaction with T4 ($\text{N}\eta 1\text{-O}2 = 2.65 \text{ \AA}$, $\text{N}\eta 2\text{-O}2 = 2.65 \text{ \AA}$). Neither the A5 nor A6 bases appear to interact with HNF-3 γ (107-223); however, the side-chain guanidinium group of Arg 214 lies in the major groove in the vicinity of these two bases and such interactions cannot be ruled out. The remaining side-chain to base contacts occur in the major groove. G7 makes water-mediated contacts with the protein (depicted schematically in Fig. 4d). In both crystallographically indepen-



dent complexes within the asymmetric unit, there is an identical network of four well ordered water molecules that mediate contacts between three amino acid side chains (Arg 214, Asn 174 and Phe 173) and G7 (O6 and N7). At the centre of this hydrogen-bond network is a very localized water molecule (refined temperature factor = $9 \text{ \AA}^2$) labelled with an asterisk in Figs 2b and 4a, b, d. Although the network surrounding G7 is somewhat more complex, it is analogous to the water network surrounding a guanine in the structure of the *trp* repressor-operator complex determined at $2.4 \text{ \AA}$ resolution by Sigler and co-workers^{28,29}. The methyl group of T8 makes a van der Waals contact with the side chain of His 169. The base of C9 does not interact with HNF-3 γ (107–223), and there are no nearby protein atoms in either groove. Finally, there is a typical two hydrogen bond contact between A10 and an asparagine side chain lying in the major groove²¹ (Asn 165(O δ 1)–A10(N6) = $3.4 \text{ \AA}$, and Asn 165(N δ 2)–A10(N7) = $3.1 \text{ \AA}$).

HNF-3 γ (107–223) makes a combination of direct and water-mediated side chain-base contacts with the sequence TAAGTCA, which lies in the middle of the HNF-3 γ -DNA footprint and corresponds to positions –105 to –99 of the TTR promoter². The three bases interacting with H3 (indicated in

bold type above) also occur within two known high-affinity HNF-3 binding sites found in the α -1 antitrypsin² and HNF-1 α ⁹ promoters. Unfortunately, simple sequence comparison studies of a number of other physiologically relevant promoters have been unable to define a consensus binding site for HNF-3 family members. The co-crystal structure determined here in combination with appropriate DNA footprinting and binding-site selection studies should help analyses of the target sequences for all of the known HNF-3/*fork head* transcription factors.

Comparison with histone H5 and HTH proteins

The structure of the unrelated, globular domain of histone H5 (GH5) determined in the absence of DNA¹³ is very similar to that of the HNF-3/*fork head* DNA-recognition motif (Fig. 5). The core histone proteins (H2A, H2B, H3 and H4) and 146 bp of DNA assemble into the nucleosome core particle. Avian erythrocyte H5 or its homologue H1 are required for further DNA packaging into a higher-order chromatin structure, known as the 30-nm filament (reviewed in ref. 30). The GH5 structure resembles HNF-3 γ (107–223), but lacks the W2 loop (Fig. 5). Alignment of equivalent residues in the two structures

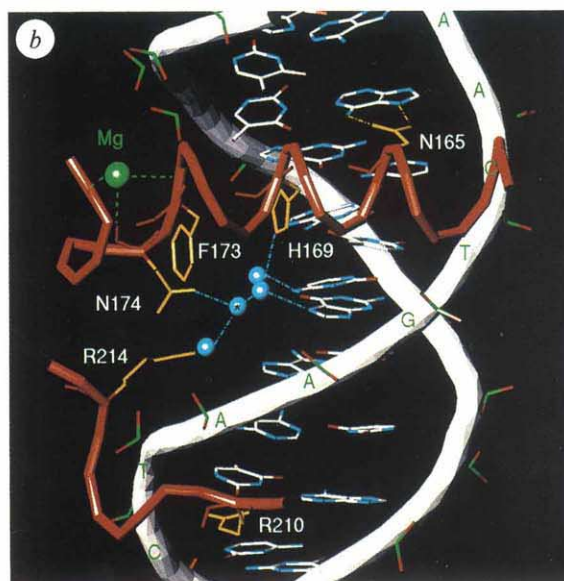
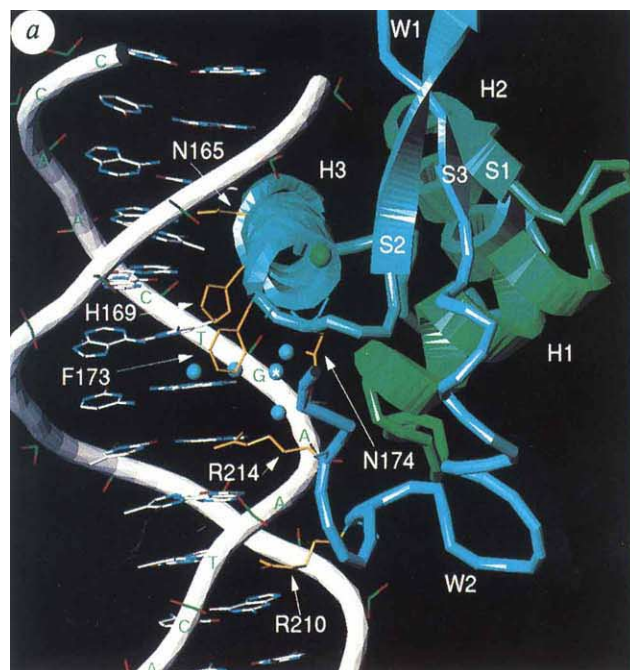
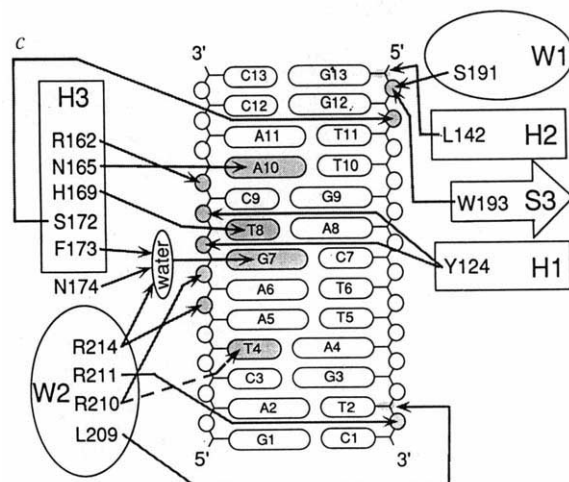
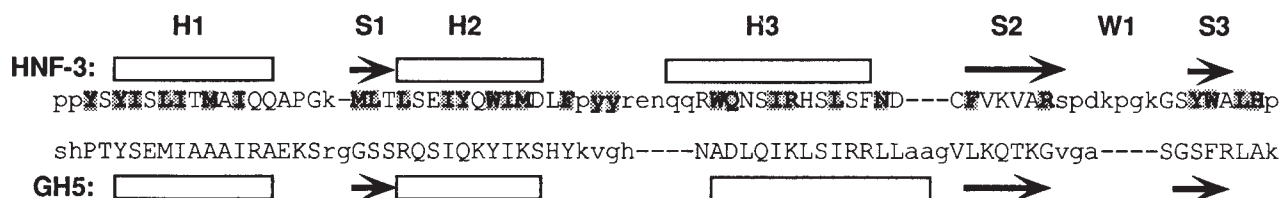


FIG. 4 HNF-3 γ (107–223)–DNA interactions. Cartoon of the side chain–base contacts, including α -helix H3 lying in the major groove making direct and H₂O-mediated interactions, and W2 straddling the phosphodiester backbone projecting the side chains of Arg 210 and Arg 214 into the minor and major grooves, respectively. The Mg²⁺ ion bound at the C terminus of H3 is depicted as a green sphere, and the water molecules are shown as blue spheres with the central water labelled with an asterisk. The base sequence of the sense strand of the TTR promoter is lettered along the path of the phosphate backbone, and the α -helices (H1, H2 and H3) and β -strands (S1, S2 and S3) are labelled. **a**, Viewed as in Fig. 3c with the polypeptide chain drawn with the colour graded from green (N terminus) to blue (C terminus). **b**, Viewed as in Fig. 3b with H3 and W2 shown in red and the remainder of the polypeptide chain removed. Putative hydrogen bonds are indicated with dashed lines. **c**, Schematic view of the HNF-3 γ (107–223)–DNA interactions. Amino acids are grouped according to their location within the protein's structural elements. Bases and phosphate groups are depicted as ovals and circles, respectively. Regions of DNA in contact with protein are shaded, and the direct and water-mediated protein–DNA contacts are indicated with arrows. The sole side chain–base minor groove contact is shown as a dashed arrow. Protein–DNA backbone contacts include Arg 214(N ϵ)–A5(P04) = $3.1 \text{ \AA}$, Arg 210(N η 2)–A6(P04) = $2.5 \text{ \AA}$, Tyr 124(N)–G7(P04) = $3.2 \text{ \AA}$, Tyr 124(OH)–T8(P04) = $2.8 \text{ \AA}$, Arg 162(N ϵ)–C9(P04) = $4.3 \text{ \AA}$, Gln 212(N ϵ 2)–C1(P04) = $4.7 \text{ \AA}$, Leu 209(C β)–T2(ribose) = $4.3 \text{ \AA}$, Ser 172(O γ)–T11(P04) = $3.0 \text{ \AA}$, Ser 191(O γ)–

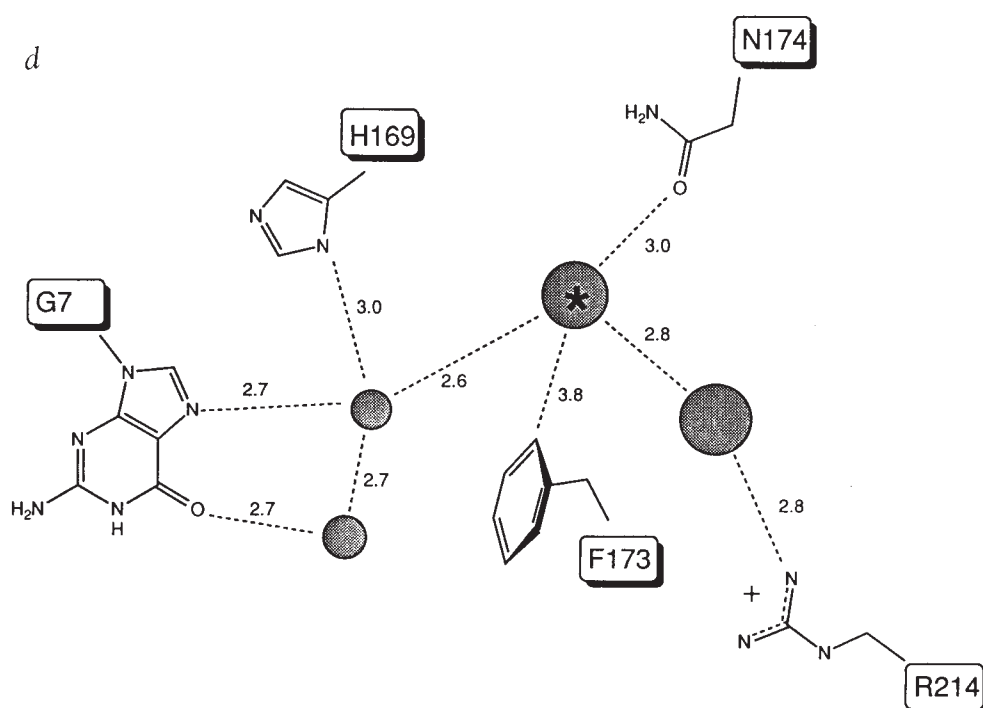


G12(P04) = $3.0 \text{ \AA}$, Trp 193(N ϵ 1)–G12(P04) = $3.0 \text{ \AA}$ and Leu 142(C δ 1)–G13(ribose) = $3.4 \text{ \AA}$. **d**, Schematic drawing of the ordered water molecules surrounding G7 in the major groove. Putative hydrogen bonds are indicated with solid lines with refined donor–acceptor distances.

FIG. 5 a, MOLSCRIPT⁴⁴ cartoons showing similarities in the following DNA-binding domains: HNF-3 γ (107–223), GH5¹³, engrailed homeo-domain (ENGR)³⁴, *Escherichia coli* catabolite gene activator protein (CAP)³³. HNF-3 γ (107–223) is viewed looking towards the DNA-binding surface as if the target DNA was running vertically. The other DNA-binding domains were drawn in the same orientation as HNF-3 γ , defined using the lsq-improve option of the molecular graphics program Q16. b, Amino-acid sequence alignment of HNF-3 γ (107–223) and GH5 based on the structural alignment (shading identifies the buried amino acids contributing to the hydrophobic core, see Fig. 1a). Secondary structural elements of each protein are indicated with rectangles for α -helices and arrows for β -strands. Structurally equivalent residues are indicated in upper case (r.m.s. deviation between equivalent α -carbon atomic positions = 1.3 Å). Similar comparisons of HNF-3 γ (107–223) with ENGR (HTH only) and CAP yielded r.m.s. deviations of 1.6 Å and 1.9 Å, respectively.

$$b$$


d



demonstrates remarkable similarity, including side chains largely buried in the protein interior (r.m.s. deviation between equivalent α -carbon atomic coordinates = 1.3 Å). The amino-acid sequence alignment derived from our structural alignment implies sequence identity of only 9 of 72 aligned residues (Fig. 5b), which is well below the 25–30% threshold for structural identity defined by Sander and Schneider²⁷.

Comparison of HNF-3 γ (107–223) with GH5 also provides some insights into the functions of both proteins. First, our protein–DNA complex structure is consistent with a mode of GH5–DNA binding predicted by Ramakrishnan *et al.*¹³. Second, the disorder exhibited by the end of loop W1 in GH5 suggests that the conformation adopted by W1 in HNF-3 γ (107–223) depends at least in part on the presence of DNA. Third, the role of the basic residues of W2 in the HNF-3 γ (107–223)–DNA interaction may be analogous to that of the highly basic C terminus of GH5, which was omitted during its crystallographic structure determination. Finally, the structural similarity of HNF-3 γ and GH5 may be functionally significant. HNF-3 binding to two adjacent, high-affinity sites in the mouse serum albumin enhancer sequence⁷ has been shown to influence phasing of the arrangement of nucleosomes within the enhancer (K. S. Zaret, personal communication), consistent with a chromatin remodelling activity of histone H5 and possibly other proteins of this structural class.

α -Helices H2 and H3 and the loop T' of the 'winged-helix' fold are structurally similar to the helix–turn–helix (HTH) proteins (reviewed in refs 21 and 32). Ramakrishnan *et al.* compared GH5 to the DNA-binding domain of *Escherichia coli* catabolite gene activator protein³³ (CAP) and estimated the r.m.s. deviation between structurally related α -carbon atoms to be 1.7 Å. Similar findings were obtained in comparisons of HNF-3 γ (107–223) versus CAP and HNF-3 γ (107–223) versus the HTH region of the engrailed (ENGR) homeodomain³⁴ (Fig. 5).

β -Turns found in canonical HTH proteins are four residues in length with glycine at the second position³⁵. If the definition of the HTH motif is loosened to permit loops instead of turns, HNF-3 γ (107–223), GH5, the third tandem repeat of c-Myb³⁶, and the Oct-1 POU-specific domain^{37,38} (POUs) can all be classified as eukaryotic HTH variants. In the c-Myb NMR structure, the 'turn' is one amino acid longer than the usual four and has

sequence Leu-Pro-Gly-Arg-Thr (see Fig. 3 in ref. 36). Overlays of the POU NMR structure on those of HTH proteins reveal a longer 'turn' (Tyr-Gly-Asn-Asp-Phe-Ser) and an extension of the C terminus of the first α -helix (see Fig. 5 in ref. 37). In GH5 and HNF-3 γ (107–223) the two 'turns' are six and eight residues long, respectively, and both the C terminus of the first and the N terminus of the second α -helices are extended relative to those in canonical HTH proteins.

Further support of our description of eukaryotic HTH variants comes from structural comparisons of two homeodomain–DNA complexes with the HNF-3 γ (107–223)–DNA complex. Like engrailed³⁴ and MAT α ³⁹, HNF-3 γ (107–223) binds B-form DNA as a monomer with the second of the 'HTH' helices lying in the major groove and other regions of the protein making a variety of base and DNA backbone contacts. It should be emphasized that earlier predictions of the DNA-binding modes of GH5, POU and c-Myb have been largely inferred by analogy with classic HTH proteins. The structure of the HNF-3 γ (107–223)–DNA complex provides the first direct proof that a eukaryotic HTH variant can bind DNA in a manner similar to the monomeric, eukaryotic HTH proteins, engrailed³⁴ and MAT α ³⁹.

In conclusion, the three-dimensional structure of this protein–nucleic acid complex explains many of the DNA-binding properties of HNF-3 γ , its homologues HNF-3 α and HNF-3 β , and the remaining members of the HNF-3/*fork head* transcription factor family. Our work also provides a starting point for further crystallographic, biochemical and genetic studies of this biologically important group of structurally related proteins. In particular, these results should help directed, systematic analyses of the mechanisms by which members of the HNF-3/*fork head* family regulate gene expression and control development in eukaryotes. Finally, the co-crystal structure of HNF-3 γ (107–223) bound to DNA provides insights into the DNA-packaging function of H1 or H5, and the more general problem of eukaryotic HTH variants.

Note added in proof: We notice that our structure of the DNA-recognition motif of HNF-3 γ is topologically identical to the unrelated, N-terminal or DNA-binding domain of BirA, the repressor of the *Escherichia coli* biotin biosynthetic operon⁴⁹. □

Received 10 March; accepted 18 June 1993.

- Lai, E., Clark, K. L., Burley, S. K. & Darnell, J. E. Jr *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Costa, R. H., Grayson, D. R. & Darnell, J. E. Jr *Molec. cell. Biol.* **9**, 1415–1425 (1989).
- Lai, E. *et al. Genes Dev.* **4**, 1427–1436 (1990).
- Lai, E., Prezioso, V. R., Tao, W., Chen, W. S. & Darnell, J. E. Jr *Genes Dev.* **5**, 416–427 (1991).
- Weigel, D., Jurgens, G., Kuttner, F., Seifert, E. & Jackle, H. *Cell* **57**, 645–658 (1989).
- Weigel, D. & Jackle, H. *Cell* **63**, 455–456 (1990).
- Liu, J.-K., DiPersio, C. M. & Zaret, K. S. *Molec. cell. Biol.* **11**, 773–784 (1991).
- Ip, Y. T., Poon, D., Stone, D., Granner, D. K. & Chalkey, R. *Molec. cell. Biol.* **10**, 3770–3781 (1990).
- Kuo, C. J. *et al. Nature* **355**, 457–461 (1992).
- Pani, L., Qian, X., Clevidence, D. & Costa, R. H., *Molec. cell. Biol.* **12**, (1992).
- Grossniklaus, U., Pearson, R. K. & Gehring, W. J. *Genes Dev.* **6**, 1030–1051 (1992).
- Dirksen, M. L. & Jamrich, M. *Genes Dev.* **6**, 599–608 (1992).
- Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L. & Sweet, R. M. *Nature* **362**, 219–223 (1993).
- Terwilliger, T. C. & Eisenberg, D. *Acta crystallogr.* **A39**, 813–817 (1983).
- Zhang, K. Y. J. & Main, P. *Acta crystallogr.* **A46**, 377–381 (1990).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. *Acta crystallogr.* **A47**, 110–119 (1991).
- Brünger, A. T. *X-PLOR v. 3.0 manual* (Yale University, New Haven, 1992).
- Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F. & Holmes, K. C. *Nature* **347**, 37–44 (1990).
- Brünger, A. T., Kuriyan, J. & Karplus, M. *Science* **235**, 458–460 (1987).
- Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. *J. molec. Biol.* **7**, 95–99 (1963).
- Pabo, C. & Sauer, R. A. *Rev. Biochem.* **61**, 1053–1095 (1992).
- Bowie, J. U., Luthy, R. & Eisenberg, D. *Science* **253**, 164–170 (1991).
- Burley, S. K. & Petsko, G. A. *Adv. Prot. Chem.* **39**, 125–189 (1988).
- Reid, K. S. C., Lindley, P. F. & Thornton, J. M. *FEBS Lett.* **190**, 209–215 (1985).
- Burley, S. K. & Petsko, G. A. *J. Am. chem. Soc.* **108**, 7995–8001 (1986).
- Li, C., Lusis, A., Sparkes, R., Tran, S. & Gaynor, R. *Genomics* **13**, 658–664 (1992).
- Sander, C. & Schneider, R. *Proteins Struct. Funct. Genet.* **9**, 56–68 (1991).
- Otwinowski, Z. S. *et al. Nature* **335**, 321–329 (1988).

- Haran, T., Joachimski, A. & Sigler, P. *EMBO J.* **11**, 3021–3030 (1992).
- Grunstein, M. A. *Rev. Cell. Biol.* **6**, 643–678 (1990).
- Crane-Robinson, C. & Pitts, O. B. *Prot. Engng* **2**, 577–582 (1989).
- Harrison, S. C. *Nature* **353**, 715–719 (1991).
- Schultz, S. C., Shields, G. C. & Steitz, T. A. *Science* **253**, 1001–1007 (1991).
- Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B. & Pabo, C. O. *Cell* **63**, 579–590 (1990).
- Brennan, R. G. & Matthews, B. W. *J. biol. Chem.* **264**, 1903–1906 (1989).
- Ogata, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **89**, 6428–6432 (1992).
- Assa-Munt, N., Mortishire-Smith, R. J., Aurora, R., Herr, W. & Wright, P. E. *Cell* **73**, 193–205 (1993).
- Dekker, N. *et al. Nature* **362**, 852–855 (1993).
- Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D. & Pabo, C. O. *Cell* **67**, 517–528 (1991).
- Tao, W. & Lai, E. *Neuron* **8**, 957–966 (1992).
- Li, C., Lai, C., Sigman, D. S. & Gaynor, R. B. *Proc. natn. Acad. Sci. U.S.A.* **88**, 7739–7743 (1991).
- Oliver, S. G. *Nature* **357**, 38–46 (1992).
- Richardson, J. S. *Meth. Enzym.* **115**, 359–390 (1985).
- Kraulis, P. J. *J. appl. Crystallogr.* **24**, 946–950 (1991).
- Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. *Meth. Enzym.* **185**, 60–89 (1990).
- Chait, B. T. & Kent, S. B. H. *Science* **257**, 1885–1894 (1992).
- Puglisi, J. D. & Tinoco, I. Jr *Meth. Enzym.* **180A**, 304–325 (1989).
- Bernstein, F. C. *et al. J. molec. Biol.* **112**, 535–542 (1977).
- Wilson, K. S. *et al. Proc. natn. Acad. Sci. U.S.A.* **89**, 9257–9261 (1992).

ACKNOWLEDGEMENTS. We thank S. Jacques for help with protein production, B. Chait and S. Cohen for the mass spectrometry at the National Resource for Mass Spectrometric Analysis of Biological Macromolecules, A. Gazes and the staff of Rockefeller University Computing Services for technical support, and V. Ramakrishnan for early transmission of the unrefined atomic coordinates of GH5. For their many useful suggestions, we thank R. Brown, A. R. Ferre-D'Amare, J. L. Kim, X. P. Kong, J. Kuriyan, J. L. Martin, D. B. Nikolov, Z. Otwinowski, G. A. Petsko and M. Rould. S.K.B. and E.L. thank J. E. Darnell Jr for his continuing support and encouragement of their collaboration. This work was supported by the Howard Hughes Medical Institute (S.K.B.) and by the American Cancer Society (E.L.).

Stellar black holes in globular clusters

S. R. Kulkarni*, Piet Hut† & Steve McMillan‡

* Division of Physics, Mathematics, and Astronomy, 105-24, California Institute of Technology, Pasadena, California 91125, USA

† Institute for Advanced Study, Princeton, New Jersey 08540, USA

‡ Department of Physics and Atmospheric Science, Drexel University, Philadelphia, Pennsylvania 19104, USA

FOLLOWING the discovery of X-ray sources in globular clusters, the accretion of matter onto a central massive black hole was suggested¹⁻³ as a possible explanation. Subsequently, it was found^{4,5} that these sources could be readily explained by thermonuclear instabilities on neutron-star surfaces and the black-hole models were abandoned. We show here, however, that the recent discovery⁶ of large populations of millisecond pulsars—and hence neutron stars—in globular clusters implies that several hundred stellar black holes (of about ten solar masses) should form within a typical cluster. In clusters of high central density, we find that the rapid dynamical evolution of the black-hole population will cause ejection of nearly all of the holes on a relatively short timescale. But in systems of intermediate density, some of the surviving holes may capture a normal star to form a low-mass X-ray binary. We suggest that there may be one or more such binaries in the globular clusters surrounding our Galaxy. These systems will be quiescent most of the time—with only occasional X-ray outbursts—but future observations of the hard X-ray spectrum may indirectly establish their existence.

Detailed dynamical studies⁷ of the rich cluster M15, perhaps the best observed globular cluster, find evidence for $4 \times 10^4 M_\odot$ (where $M_\odot$ is the mass of the Sun) of dark matter. This dark matter is attributed to massive white dwarfs, the stellar remnants of main-sequence stars with mass $M \gtrsim 3 M_\odot$, and neutron stars, the remnants of stars with mass $M > M_N = 8 M_\odot$. Phinney⁸ finds that the data are well fitted with a Salpeter initial mass function (IMF), $dN/dM \propto M^{-\alpha}$ with $\alpha \approx 2.35$. The predicted number of neutron stars at early times is $N_n \sim 10^4$. It is an observed fact, however, that neutron stars tend to have large velocities at formation, so we expect only a fraction, $f_N \sim 0.3$, to be retained by the shallow gravitational potential well of a typical globular cluster⁹.

Such a large estimate for the present number of neutron stars in clusters, N_n is consistent with the huge abundance of cluster pulsars⁶. Moreover, the radio pulsar observations also show that N_n depends only on gross structural parameters (total mass, core density). No apparent dependence has been found either on the metallicity of the cluster (for example M15, in which at least eight pulsars¹⁰ have been found, is metal poor, whereas 47 Tuc, with at least ten known pulsars¹¹, is metal-rich), or on the value of α for the currently visible stars with $M \lesssim 0.8 M_\odot$. Pulsars have also been found in moderate- and low-density clusters such as M13 and M53.

We assume that the cluster pulsars are neutron stars, the stellar remnants of massive stars, spun up by accretion of matter from a companion (that is during the low-mass X-ray binary (LMXB) phase)—the usual model for millisecond pulsars. Others¹² argue that the Salpeter IMF will result in cluster disruption (due to rapid mass loss from stellar evolution), and have speculated that all millisecond pulsars are accretion-induced transmuted white dwarfs born with low velocities. The mass of such neutron stars, however, would be $1.25 M_\odot$, whereas the best-measured cluster neutron stars have masses similar¹³ to 'ordinary' pulsars. In addition, the scale height of the observed millisecond pulsars in the disk of the Galaxy is consistent with large space motions for such objects. We therefore favour the standard model, and now consider the logical extension of this hypothesis—black holes, the stellar remnants of even more massive stars, with $M > M_B$.

The value of M_B , unlike M_N , is not well known. Evolutionary models and statistics of Galactic black-hole systems suggest¹⁴ $M_B \sim 60 \pm 20 M_\odot$. Models¹⁵ of chemical evolution of galaxies at early times, and hence with metallicities representative of proto-globular-clusters, require stars above $25 M_\odot$ to collapse to black holes without ejecting processed material. The precise mass of the resulting black holes, m_b , depends upon the mass-loss history. For the Galactic binaries, we find estimates of m_b to be $16 M_\odot$ (Cyg X-1; ref. 16), $12 M_\odot$ (GS 2023+338; ref. 17), $6 M_\odot$ (Nova Muscae; ref. 18), $7 M_\odot$ (A0620-00; ref. 16) and $7 M_\odot$ (LMC X-3; ref. 16). We adopt $m_b = 10 M_\odot$. This assumption contrasts with an earlier study¹⁹ which considered the structural influence of a substantial hypothetical population of low-mass ($\lesssim 2.5 M_\odot$) black holes.

In the absence of any compelling evidence to the contrary, we make the reasonable assumption that the IMF for stars with $M \geq M_N$ follows a Salpeter distribution, as appears to be the case for M15. With this assumption, and starting with a conservative estimate $N'_n \sim 3 \times 10^3$, we find $N_b \sim 150-400$, for the above range in M_B . Unlike the neutron stars, the black-hole X-ray binaries in the Galactic disk do not show peculiar radial motion as large as LMXBs (S.R.K., manuscript in preparation). Consequently, we assume that the retention factor $f_B \lesssim 1$. We will start with the conservative assumption that rich clusters began with $N_b \sim 10^2$ black holes of mass $m_b = 10 M_\odot$, and consider their fate as the cluster moves inexorably toward core collapse^{20,21}. Later we will indicate how the evolution of the black-hole subsystem changes as N_b increases.

Once formed, the black holes soon come into energy equipartition with the typical star (of mass $m \sim 1 M_\odot$), slow down, and settle to the centre of the cluster. This process of mass segregation takes place on the dynamical-friction time scale, $\sim t_{RH}/\mu \sim 0.1 t_{RH}$, where $\mu = m_b/m$, and t_{RH} ($\sim 10^9$ yr typically) is the cluster half-mass relaxation (thermalization) time²⁰. The process is complete by the time the cluster is $\sim 10^9$ yr old, and most of the holes are in equipartition in the core. Because the holes constitute only a small fraction of the total mass of the cluster, they do not greatly affect its overall dynamical evolution. Equipartition in the harmonic potential determined by the core stars gives $r_b^2/r_c^2 \sim v_b^2/v_c^2$, where r_b and v_b are the half-mass radius (that is, the radius enclosing the inner half of the cluster mass) and velocity dispersion, respectively, of the black holes, and r_c and v_c are core radius and velocity dispersion of the stars. Thus the holes become concentrated in the core, with $r_b/r_c \sim \mu^{-1/2} \sim 1/3$.

The long-term evolution of the system is driven by two-body stellar encounters, which allow individual stars to exchange energy (on a relaxation time scale), and provide a mechanism for heat conduction from the core of the cluster to its periphery. As a result, the core collapses—both r_c and N_c , the number of stars in the core, decrease (formally) to zero in $\sim 10 t_{RH}$ (ref. 20). During this process, the central regions can be reasonably approximated as an isothermal sphere, and the central stellar mass density increases roughly as $\rho_c \propto r_c^{-2} \propto N_c^{-2}$. However, the central mass density of the holes increases faster, $\propto r_b^{-3} \propto r_c^{-3}$, so long as the number of holes is conserved. The ratio of the central density of the holes to that of the stars is

$$\rho_b/\rho_c \sim \mu (N_b/N_c) (r_c/r_b)^3 \sim \mu^{5/2} (N_b/N_c)$$

Therefore, if the holes did not interact with one another, their central density would become comparable to that of the stars by the time the core had shrunk to contain $N_c \sim N_b \mu^{5/2} \sim 3 \times 10^4$ stars. At around this time, the Spitzer mass-segregation instability²⁰—the runaway collapse of the black-hole subsystem when it decouples from the stellar core—would be expected to set in.

The holes do interact with each other however, forming hole-hole binaries which subsequently eject single holes from the cluster. The timescale for 'three-body' hole-hole binary-formation (in which a third hole carries away the excess energy required

for two others to become bound) is²²

$$t_{3B} \sim 3\mu^{-25/3} N_b^{-3} N_c^5 \tau_c$$

where $\tau_c = r_c^{3/2}/(GN_c m)$ is the time required for a typical star to traverse the core. We assume that this process starts to become important when $t_{3B} \sim 0.1 t_{cc} \sim 5 t_{Rc}$ (the coefficient depending somewhat on the evolutionary state of the core), with t_{cc} the time left until core collapse (that is, when $r_c = 0$) and t_{Rc} the two-body relaxation time for the core stars:

$$t_{Rc} \sim 0.1 N_c \tau_c$$

In this expression, we have taken the usual 'Coulomb' logarithmic factor²⁰ to be ~ 10 . Equating t_{3B} and $0.1 t_{cc}$, we obtain:

$$N_c \sim 0.6 \mu^{25/8} N_b^{3/4}$$

For our choice of parameters, we again find $N_c \sim 3 \times 10^4$. For a typical globular cluster, this corresponds to $r_c \sim 1$ pc and $t_{Rc} \sim 2 \times 10^8$ yr. Thus hole-hole binaries begin to form several billion years before core collapse, effectively preventing development of the mass-segregation instability.

For larger values of N_b , both the onset of the mass-segregation instability and the formation of the first black-hole binary occur earlier in the cluster evolution. From the above scaling relations, however, we see that the Spitzer instability is (slightly) favoured as N_b increases. Once this instability sets in, the black-hole subsystem decouples from the stellar core and enters a phase of rapid dynamical evolution, undergoing core collapse on a timescale $\sim t_{Rb} \sim \mu^{-5} (N_c/N_b) t_{Rc} \sim \mu^{-5/2} t_{Rc} < t_{cc}$. The resulting high spatial density leads to the almost immediate formation of black-hole binaries once runaway mass segregation begins. The production of binaries continues until N_b falls below the threshold for stability. The remaining few holes must wait for the core to shrink to the point where the previous discussion applies. In our specific estimates below, we will use our initial assumption of $N_b \sim 10^2$. Also, we will focus on stellar-mass black holes, that is, we will not discuss the possibility that a post-collapse runaway merging of normal stars will produce a single massive black hole²³.

Once formed, a black-hole binary becomes more tightly bound ('hardens') as a result of superelastic interactions with other holes²⁴. Its recoil velocity after each encounter increases, and it is eventually ejected from the cluster. Following Hut *et al.*²⁵, and taking into account the greater mass of the holes, we estimate the time required for a hole-hole binary to escape from the cluster in this manner to be $\sim 3 \times 10^3 t_{Rb} \sim 10 t_{Rc}$ (a few times 10^8 yr). Based on N -body simulations of binaries in small- N systems (here, the black-hole subsystem), we expect that at most one or two hole-hole binaries will be present in the cluster at any given time. As each binary ejects ~ 5 – 10 single black holes as it hardens, we conclude that the black-hole population will be depleted in less than a billion years, well before core collapse is complete. From the previous discussion, we see that this conclusion is largely independent of the precise initial value of N_b .

We note that superelastic encounters between black-hole binaries and stars may inject a substantial amount of energy into the cluster. The total energy supplied by a single black-hole binary as it hardens from formation at ~ 1 kT to ejection at $\sim 10^4$ kT (where $\frac{3}{2} kT$ is the mean stellar kinetic energy) is $\sim 3 \times 10^3$ kT, or about 5–10% of the core kinetic energy. The bulk of this energy is actually spread over a fairly large fraction of the entire cluster, as reaction products recoil into the halo and the black holes slowly drift back to the centre by dynamical friction. We therefore expect that, while binary heating may significantly slow the overall collapse of the stellar core, it will probably not halt it completely, at least for $N_b \leq 10^3$. In any case, these hole-star interactions will have little direct influence on the black-hole subsystem itself, as it evolves primarily by hole-hole encounters.

By the end of the core-collapse phase, it is likely that the last hole-hole binary will still remain in the core, having ejected all

other holes. There is a small chance that this final binary will itself have escaped following the interaction that ejected the last single hole. Similarly, there is a small chance that a lone third hole is left on a very wide 'parking' orbit, with its apocenter at several half-mass radii²⁵. The ejected black-hole binaries will have orbital separations of ~ 1 AU. The timescale on which they are expected to coalesce by means of the emission of gravitational radiation is $\sim 10^{14}$ yr, much longer than the age of the Universe. Thus we expect that they will not be significant sources for future gravitational wave interferometers.

We have seen that there is an interval of $\sim 10^9$ yr, probably occurring well before core collapse, when black-hole binaries can form and the hole population is largely depleted. During this period, a few holes may acquire companions through tidal capture, forming close binaries with typical separations of a few (5–10) stellar radii. At early times, the stellar density would not be high enough for this process to be likely; at late times the holes will be ejected from the core (although this will probably leave the hole-star binary intact, as the hole-hole binary will have a size $\sim 10^2$ larger). Although the existence of these binaries is unimportant to the overall dynamics of the cluster, they do provide us with a possible, (if transient), means of observing the black-hole subsystem. The number of black-hole tidal-capture binaries formed in our model cluster is $\sim 0.01 N_b$. Given the measured rate of core collapse²¹, we estimate that there are currently $2 \times 0.01 N_b \sim 10$ such binaries in the Galactic globular cluster system.

Mass transfer begins only when the companion evolves onto the subgiant branch. The probability that this will take place for a typical core star during the $\sim 10^9$ yr before the black hole is ejected from the cluster is ~ 0.1 . Therefore, we estimate that the number of active hole-containing low-mass X-ray binaries in the globular cluster system is ~ 1 . Note that our analysis also predicts the presence of several such black-hole binaries in the Galactic bulge and halo, as a result of ejection a few billion years ago.

Like their Galactic disk counterparts (for example A0620-00, ref. 26), the cluster accreting black-hole binaries are expected to be transients. An explanation²⁷ for this transient behaviour is an accretion disk instability that arises only when $q \equiv M_d/M_a < 0.25$, where M_d and M_a are the masses of the donor and accretor stars. This condition is easily satisfied by the proposed cluster black-hole binaries. The disk black-hole binary transients show hard X-ray tails and it would be desirably to search for (and then monitor) hard X-ray transients with GRANAT, ASCA, the Compton Gamma Ray Observatory and future missions like the X-ray Timing Explorer (especially the All Sky Monitor).

Recently, Brown and Bethe²⁸ have suggested that main-sequence stars with $18 M_\odot < M < 25 M_\odot$ evolve to low-mass ($1.5 M_\odot$) black holes, and that stars above $25 M_\odot$ evolve to high-mass ($\geq 10 M_\odot$) black holes of the sort seen in the Galactic disk. In clusters, such low-mass black holes behave dynamically like neutron stars. Their number is $N_{imb} \sim N_n f_{imb}$, where f_{imb} is the unknown retention factor. Because all cluster LMXBs²⁹ (save two for which few observations exist) show bursts, we conclude that either $f_{imb} \ll 1$, or that this mechanism does not operate.

Holes in clusters offer the possibility of a black-hole-millisecond pulsar binary system with large eccentricity and small orbital period, a system ordinarily not possible in the Galactic disk. Such systems can be profitably used for tests of general relativity (akin to studies of binary pulsar systems³⁰). Radio pulsar searches that are sensitive to compact binaries, although time consuming, may prove fruitful in the long run. $\square$

Received 3 March; accepted 8 June 1993.

1. Bahcall, J. N. & Ostriker, J. P. *Nature* **256**, 23–24 (1975).
2. Silk, J. & Arons, J. *Astrophys. J.* **200**, L131–L135 (1975).
3. Grindlay, J. E. *Astrophys. J.* **221**, 234–257 (1978).
4. Grindlay, J. E., Hertz, P., Steiner, J. E., Murray, S. S. & Lightman, A. P. *Astrophys. J.* **282**, L13–L16 (1984).
5. Joss, P. C. & Rappaport, S. A. *A. Rev. Astr. Astrophys.* **22**, 537–592 (1984).
6. Johnston, H. M., Kulkarni, S. R. & Phinney, E. S. in *X-ray Binaries and Recycled Pulsars*

- NATO: ASI Series (eds van den Heuvel, E. P. J. & Rappaport, S. A.) 349–364 (Kluwer, Dordrecht, 1992).
7. Grabhorn, R. P., Lugger, P. M. & Murphy, B. W. *Astrophys. J.* **392**, 86–98 (1992).
 8. Phinney, E. S. *Phil. Trans. R. Soc. A* **341**, 39–75 (1992).
 9. Hut, P., Murphy, B. W. & Verbunt, F. *Astr. Astrophys.* **241**, 137–141 (1991).
 10. Anderson, S. B. thesis, California Institute of Technology (1992).
 11. Manchester, R. N. et al. *Nature* **352**, 219–221 (1991).
 12. Bailyn, C. D. & Grindlay, J. E. *Astrophys. J.* **353**, 159–167 (1990).
 13. Prince, T. A., Anderson, S. B., Kulkarni, S. R. & Wolszczan, A. *Astrophys. J.* **372**, L75–L78 (1991).
 14. van den Heuvel, E. P. J. in *Proc. European ISY 1992 Conference* (ESA Publication Division: ESTEC Noordwijk, in the press).
 15. Maeder, A. *Astr. Astrophys.* **264**, 105–120 (1992).
 16. Cowley, A. P. A. *Rev. Astr. Astrophys.* **30**, 287–310 (1992).
 17. Casares, J., Charles, P. A. & Naylor, T. *Nature* **355**, 614–617 (1992).
 18. Remillard, R. A., McClintock, J. E. & Bailyn, C. D. *Astrophys. J.* **399**, L145–L149 (1992).
 19. Larson, R. B. *Mon. Not. R. astr. Soc.* **210**, 763–777 (1984).
 20. Spitzer, L. *Dynamical Evolution of Globular Clusters*, 51–71 (Princeton Univ. Press, 1987).
 21. Hut, P. & Djorgovski, S. *Nature* **359**, 806–808 (1992).
 22. Goodman, J. & Hut, P. *Astrophys. J.* **403**, 271–277 (1993).
 23. Lee, H. M. in *Dynamics of Dense Stellar Systems* (ed. Merritt, D.) 105–112 (Cambridge Univ. Press, 1989).
 24. Heggie, D. C. *Mon. Not. R. astr. Soc.* **173**, 729–787 (1975).
 25. Hut, P., McMillan, S. & Romani, R. *Astrophys. J.* **389**, 527–545 (1992).
 26. McClintock, J. E. & Remillard, R. A. *Astrophys. J.* **308**, 110–122 (1986).
 27. Whitehurst, R. & King, A. *Mon. Not. R. astr. Soc.* **249**, 25–35 (1991).
 28. Brown, G. E. & Bethe, H. A. *Astrophys. J.* (submitted).
 29. Verbunt, F., Hasinger, G., Johnston, H. M. & Bunk, W. Univ. of Utrecht preprint (1992).
 30. Taylor, J. H. *Phil. Trans. R. Soc. A* **341**, 117–134 (1992).

ACKNOWLEDGEMENTS. We thank L. Bildsten, M. Davies, S. Thorsett and F. Verbunt for discussions. S.R.K. was supported by NSF, NASA and the Packard Foundation. This work was supported by grants from NASA (S.M.) and the Pittsburgh Supercomputer Center (P.H. and S.M.).

Primordial black holes in globular clusters

Steinn Sigurdsson & Lars Hernquist

Board of Studies in Astronomy and Astrophysics,
University of California, Santa Cruz, California 95064, USA

IT HAS recently been recognized¹ that significant numbers of medium-mass black holes (of order 10 solar masses) should form in globular clusters during the early stages of their evolution. Here we explore the dynamical and observational consequences of the presence of such a primordial black-hole population in a globular cluster. The holes initially segregate to the cluster cores, where they form binary and multiple black-hole systems. The subsequent dynamical evolution of the black-hole population ejects most of the holes on a relatively short timescale: a typical cluster will retain between zero and four black holes in its core, and possibly a few black holes in its halo. The presence of binary, triple and quadruple black-hole systems in cluster cores will disrupt main-sequence and giant stellar binaries; this may account for the observed² anomalies in the distribution of binaries in globular clusters. Furthermore, tidal interactions between a multiple black-hole system and a red-giant star can remove much of the red giant's stellar envelope, which may explain the puzzling absence³ of larger red giants in the cores of some very dense clusters.

Theories of the evolution of massive stars suggest that some fraction of the most massive ones will form black holes of ~ 5 – $20 M_{\odot}$ (ref. 4), where $M_{\odot}$ is the mass of the Sun. The end product of stellar evolution, whether a neutron star, black hole, or complete disruption of the star, may depend on the rotation of the stellar core and the metallicity of the envelope, hence the fraction of stars that collapse to black holes is not well determined. Models suggest that low metallicity stars may form black holes more readily than stars of solar metallicity⁵. For a range of initial mass functions (IMFs) and stellar masses, estimates suggest that from 10^{-4} to less than 10^{-6} of stars will end as black holes. The same process of stellar evolution should produce black holes in globular clusters⁶ (S. Kulkarni, personal commu-

nication) with interesting observational and dynamical implications. As globular clusters contain of order 10^6 stars we expect between zero and 100 black holes to form in each cluster. If the IMF is biased towards massive stars at high densities, the fraction of stars leading to black holes may be greater in protocluster cores¹. Globulars contain a large number of neutron stars⁷. Assuming that most of these neutron stars are 'primordial' and not the product of accretion induced collapse⁸, and as the star formation process is unlikely to 'know' the mass at which evolved stars collapse to black holes (rather than neutron stars) we can have some confidence that the IMF does not steepen sharply at that critical mass, and that black holes are likely to form in protoclusters. Because black holes, unlike neutron stars, are not thought to receive kicks when they are produced⁹, and recoil from binary dissociation during black hole formation will be small, primordial black holes ought to be retained in globulars when formed. A dynamically relaxed cluster forms an equilibrium 'core', where the stellar density gradient is small and stars are in kinetic equilibrium, moving with a velocity dispersion $\sim \sigma \times (\bar{m}/m_*)^{1/2}$, where m_* is the mass of a star, moving in a multi-mass field with average stellar mass $\bar{m}$. Typically, $\sigma \sim 5 \text{ km s}^{-1}$. Black holes will sink to the core on a dynamical friction timescale, t_{df}

$$t_{\text{df}} \sim 3 \times 10^7 \left(\frac{\sigma}{10 \text{ km s}^{-1}} \right) \left(\frac{r_c}{1 \text{ pc}} \right) \left(\frac{M_{\text{BH}}}{10 M_{\odot}} \right)^{-1} \text{ yr} \quad (1)$$

where r_c is the cluster core radius and M_{BH} is the mass of the black hole. This process starts after the stellar population has evolved to the point where the main-sequence turnoff is well below $5 M_{\odot}$, and all primordial black holes and neutron stars have formed. Typically this will be a few hundred million years after the formation of the cluster.

In the core, the black holes will form binaries. Stars with masses 5 – $1.4 M_{\odot}$ evolve to white dwarfs in a few billion years, when any primordial black holes will be an order of magnitude more massive than the surviving stars. The black holes random-walk (by stellar 'brownian motion') in the core over a radius of order 10^5 AU . Collective effects lead to the black holes becoming bound into black hole–black hole (BHBH) binaries¹⁰. The binding energy of BHBH binaries exceeds the stellar kinetic energy at semi-major axis ($a \lesssim 10^3 \text{ AU} (M_{\text{BH}}/10 M_{\odot})^2 / (\sigma/10 \text{ km s}^{-1})^2$) and the black hole binaries are hardened by encounters with stars and other black holes; that is a shrinks and the binary binding energy increases. If the number of black holes in the core is large (>4), the black holes and BHBH binaries interact. A strong BH–BHBH encounter releases about 1/3 of the BHBH binding energy¹¹. This energy is released as kinetic energy of the bodies, with each component recoiling at a speed inversely proportional to its mass. A BHBH binary can recoil from the core once the binaries have hardened to semi-major axis of $\sim 30 \text{ AU}$. Black holes and BHBH binaries ejected from the core sink back because of dynamical friction and continue interacting. At $\lesssim 10 \text{ AU}$, BH–BHBH encounters eject the black holes from the cluster. This process occurs on a timescale $\sim 10^8 \text{ yr} / [(a/10^3 \text{ AU})(n_{\text{BH}}/10^4 \text{ pc}^{-3})]$ (refs 11, 12), where n_{BH} is the number density of black holes in the core. (Because dynamical friction will bring the black holes into the inner 0.3 pc in all but the lowest density globulars, $n_{\text{BH}} \geq 10^2$ after the initial relaxation of the black-hole population in a young cluster rich in black holes.) Independent of the initial number we expect most of the black holes to be ejected from the cluster on a timescale short compared to the cluster lifetime. The final black-hole population will depend on the random encounters of the last one or two black-hole binaries and whether these provide sufficient recoil to eject the black holes from the cluster, or whether the black holes end up on wide orbits beyond the cluster half-mass radius but bound to the cluster. We do not expect all clusters to contain black holes at this stage. In order to estimate the fraction of clusters that eject all their black holes, it will be necessary to perform

Monte Carlo simulations of the small numbers of black holes diffusing into cluster cores and undergoing strong interactions leading to ejections. Analytic estimates provide poor constraints as the escape velocity of globulars is not well determined.

Some clusters will retain a few black holes which will be of two types. First, those formed in the halo of the cluster and which had not reached the core before the frenzy of interaction ejected most of the black hole population. Second, those ejected onto wide halo orbits, and circularized by orbit diffusion at radii $r \sim (r_c r_e)^{1/2}$, (where r_c is the cluster core radius and r_e is the radius to which the black hole was ejected) before returning to the core on a relaxation timescale $t_r \sim 2 \times 10^8 (r/5 \text{ pc})^{3/2} / (M_{\text{BH}}/10 M_\odot) \text{ yr}$. A black hole (binary) entering the core (now depleted of black holes), will sink to the core through dynamical friction; if another black hole (binary) enters the core it will become bound to the first black hole (binary) on a timescale t_{df} . If one or both are binaries, there is a probability of order a/r_c that there will be an immediate encounter leading to recoil of the black holes from the core. Alternatively the system will form a wide binary (if one or both were already binaries, a hierarchical triple or quadruple system will form), with $a \sim 10^3 \text{ AU}$ as before. At this point the dynamical evolution of the core black hole population slows down. The system now slowly hardens through stellar encounters at a rate $\sim 1/t_h$, where $t_h = 10^8 \text{ yr} / [(a/10^3 \text{ AU})(n_4)]$ and $n_4 = n/10^4 \text{ pc}^{-3}$ is the core stellar density¹⁰. This process continues either until $a \sim 10 \text{ AU}$ and there is a triple black hole interaction with ejection, or, if only two holes remain, $a \sim 0.1 \text{ AU}$, at which point stellar encounters provide sufficient recoil to eject the black hole binary. With $m_* \ll M_{\text{BH}}$, efficiency of binding energy extraction is reduced¹² by $(M_*/M_{\text{BH}})^{1/2}$. Massive-field stars, such as neutron stars with mass m_{NS} , are disproportionately efficient in ejecting the black hole binary, if their number fraction in the core is no less than $\bar{n}/m_{\text{NS}}$. The density of massive stars in the core scales like $N_0 \bar{n}(r)^{m_{\text{NS}}/m_*}$, where $\bar{n}(r)$ is the mean density; the normalization constant $N_0 \lesssim 5$ is somewhat constrained by observed cluster mass to light ratios. Low mass stars also increase the mass to light ratio and preclude better than order of magnitude estimates of N_0 , except possibly in M15 (ref. 13). If ejection through stellar encounters is avoided, gravitational radiation leads to spiral-in and merger of the BHBH binary at $a \sim 0.03 \text{ AU}$. Hardening to the point where gravitational radiation leads to merger in less than a Hubble time requires $n_4 \gtrsim 10^2$. Such stellar densities only occur in core-collapsed globulars, of which there are about 20 in the Galaxy¹⁴. For unequal mass BHBH binaries, asymmetric emission of gravitational radiation may carry off enough linear momentum for the merged remnant to recoil from the core. Recent calculations¹⁵ however, suggest maximum recoil is of order 10 km s^{-1} which is insufficient for ejection. The rate for BHBH binary mergers in globulars is no more than 10^{-9} yr^{-1} in the galaxy. Within 100 Mpc of Earth, which is close enough for the gravitational radiation amplitude at Earth to be in the expected sensitivity range for the Laser Interferometer Gravitational Wave Observatory¹⁶; this equates to a rate of approximately one per decade.

A hierarchical triple or quadruple system will transfer $\sim 10^{47} \text{ ergs}/(a/10 \text{ AU})$ to the core, this is small compared to the core binding energy of $10^{49}/(N_*/10^4) \text{ ergs}$, where N_* is the number of stars in the cluster core. In the case where only two black holes remain in the core, hardening by stellar encounters to 0.1 AU releases $\sim 10^{49} \text{ ergs}$, but most of that energy is deposited in single stars recoiling at high velocity, which escape the core before transferring their energy. The loss of mass from the core 'heats' the core, expanding the core radius and lowering the core density, but the process is not efficient in delaying core collapse.

Differences in environment and dynamical influences lead to qualitatively different detection scenarios for black holes in globulars compared to galaxies¹⁷. In globular clusters there is no intra-cluster gas to fuel an accretion disk around black holes. Tidal encounters between single black holes and field stars occur

on a timescale of $\sim 10^{10}/n_4 \text{ yr}$. Due to resonant interactions, tidal interactions between BHBH binaries and field stars are expected to be up to two orders of magnitude more common and may be of observational interest^{6,11,12}. Currently observed X-ray sources in globulars are consistent with being neutron star binaries¹⁸. Wide BHBH binaries may not follow a relaxed distribution in the cluster core during an X-ray luminous accretion phase, as both the tidal interaction causing the accretion phase, and subsequent perturbative interactions between field stars and the binary, 'kick' the binary away from the core, and the binary may never fully relax to the core¹². A tidal interaction or physical collision between a star and a BHBH binary generally increases the angular momentum of the binary by $\Delta J/J \sim 1\%$ ($\sigma/5 \text{ km s}^{-1}/(a/10 \text{ AU})^{1/2}$), as the trajectory of the star will not be coplanar with the binary. Collisions soften (a increases) the BHBH binary on average, extending the phase of stellar binary disruption.

Stars ejected from the core by BHBH binaries induce velocity anisotropy of a few percent in the cluster, whereas the presence of black hole (binaries) in the core efficiently destroys box orbits in the cluster¹⁹, leading to spatial isotropy. This is consistent with observations of low ellipticity and significant velocity anisotropy in globulars²⁰. A point mass at the core of a globular induces a cusp both in the radial density profile of the cluster and the velocity dispersion¹. Because of the flattening of the luminous core profile in the presence of a massive dark population, observations of core density profiles are a relatively weak constraint on core dark matter³¹. Current limits on cusps in globulars require the globular to have black holes with a total mass of $< 100 M_\odot$ in their cores^{22,23,13}, consistent with our estimate.

Black holes in globulars have a strong effect on the core population of stellar binaries. Searches for main-sequence and giant binaries in globulars suggest that the number fraction of main-sequence binaries in some clusters may be significantly lower than in the Galaxy²⁰. Wide binaries in clusters are destroyed by encounters with field stars, and by encounters with other binaries^{24,25}. Exchange interactions and tidal interactions also dynamically modify the binary population, particularly in the denser clusters^{26,12}. New observations show that the number fraction of turnoff-mass binaries in the core of M3 may be significantly lower than either in the halo of M3 or in the core of the low density cluster NGC 5053 (ref. 2). With a core density of $\sim 10^{3.5} \text{ pc}^{-3}$, the density of M3 is not sufficient for the binary population to have been significantly disrupted by stellar binary interactions. A binary black hole, or better still, a hierarchical triple or quadruple black hole in the core of a globular provide a very efficient mechanism for destroying stellar binaries in cores. The destruction can either involve direct ejection of both members of the stellar binary during a close encounter with the black holes, or ejection of one of the stars with the other going into a (quasi)stable orbit about one of the black holes. The timescale for binary disruption, t_b , for a binary fraction f_b , is

$$t_b \approx \frac{10^4 \text{ yr}}{(a/10^3 \text{ AU}) n_4 f_b} \quad (2)$$

A black hole multiple can destroy of order 10^3 stellar binaries before reaching $a \sim 10 \text{ AU}$, at which point recoil is likely to remove the black holes from the core and the interaction rate becomes small. Clusters with 2–4 black holes bound should efficiently destroy their core binary population in an episodic manner. The core binaries are then replenished from the cluster halo on timescale t_r , after the black hole multiple is hardened to ejection. This process should produce anomalous binary population gradients in some of the galactic globular clusters and a consequent anomalous gradient in the exotic stellar products of binaries, in particular blue stragglers.

Recent observations suggest that the tip of the red-giant branch in some very dense globular cores³, that is the largest red giants are missing from the core. Black hole hierarchical multi-

ples can tidally strip the envelopes of red giants with high probability as the radius of the giant envelope, R_* , increases. The fractional cross-section for close approach (b) to a member of a black hole multiple for a turnoff-mass giant scales like $(b/a)^\gamma$, with $\gamma \lesssim 0.5$ (ref. 12). The timescale for tidal interactions that could strip the envelope of the giant ($b \sim 3R_*$, ref. 27) becomes comparable with the giant evolution timescale at $n_4 \geq 10^2$, for a black hole multiple with $a \geq 10$ AU. A hierarchical black hole system in a collapsed cluster core can efficiently truncate the tip of the giant branch accounting for the puzzling colour gradient observed. $\square$

Received 3 March; accepted 18 June 1993.

1. Larson, R. B. *Mon. Not. R. astr. Soc.* **210**, 763–777 (1984).
2. Bolte, M., Hesser, J. E. & Stetson, P. B. *Astrophys. J.* **408**, L89–L92 (1993).
3. Djorgovski, S. G., Piotto, G., Phinney, E. S. & Chernoff, D. F. *Astrophys. J.* **372**, L41–L44 (1991).
4. Woosley, S. E. & Weaver, T. A. *Rev. Astr. Astrophys.* **24**, 205–253 (1986).
5. Maeder, A. *Astr. Astrophys.* **264**, 105–120 (1992).
6. Kulkarni, S. R., Hut, P. & McMillan, S. *Nature* **364**, 421–423 (1993).
7. Phinney, E. S. *Phil. Trans. R. Soc. A341*, 39–76 (1992).
8. Grindlay, J. E. & Bailyn, C. D. *Nature* **336**, 48–50 (1988).
9. Kulkarni, S. R. *Phil. Trans. R. Soc. A341*, 77–92 (1992).
10. Begelman, M. C., Blandford, R. D. & Rees, M. J. *Nature* **287**, 307–309 (1980).

11. Hut, P. & Bahcall, J. *Astrophys. J.* **268**, 319–341 (1983).
12. Sigurdsson, S. & Phinney, E. S. *Astrophys. J.* (in the press).
13. Phinney, E. S. in *Dynamics of Globular Clusters* (eds Djorgovski, S. G. & Meylan, G.) 141–169 (PASP, San Francisco, 1993).
14. Trager, S. C., Djorgovski, S. G. & King, I. R. in *Dynamics of Globular Clusters* (eds Djorgovski, S. G. & Meylan, G.) 347–355 (PASP, San Francisco, 1993).
15. Wiseman, A. G. *Phys. Rev. D46*, 1517–1539 (1992).
16. Thorne, K. S. in *300 Years of Gravitation* (eds Hawking, S. W. & Israel, W.) 330–458 (Cambridge Univ. Press, 1987).
17. Hut, P. & Rees, M. J. *Mon. Not. R. astr. Soc.* **259**, 27P–30P (1992).
18. Grindlay, J. E., Hertz, P., Steiner, J. E., Murray, S. S. & Lightman, A. P. *Astrophys. J.* **282**, L13–L16 (1984).
19. Gerhard, O. E. & Binney, J. *Mon. Not. R. astr. Soc.* **216**, 467–502 (1985).
20. Gunn, J. E. & Griffin, R. F. *Astr. J.* **84**, 752–773 (1979).
21. Grubhorn, R. P., Cohn, H. N., Lugger, P. M. & Murphy, B. W. *Astrophys. J.* **392**, 86–98 (1992).
22. Lauer, T. R. et al. *Astrophys. J.* **369**, L45–L49 (1991).
23. Shapiro, S. L. in *Dynamics of Star Clusters* (eds Goodman, J. & Hut, P.) 373–413 (Reidel, Dordrecht, 1985).
24. Hut, P., McMillan, S. & Romani, R. W. *Astrophys. J.* **389**, 527–545 (1992).
25. Leonard, P. J. T. *Astr. J.* **98**, 217–226 (1989).
26. Phinney, E. S. & Sigurdsson, S. *Nature* **349**, 220–223 (1991).
27. Hills, J. G. *Astr. J.* **102**, 704–715 (1991).

ACKNOWLEDGEMENTS. We thank S. Kulkarni, G. Quinlan and J. Ostriker for helpful comments and discussion, and S. Kulkarni for drawing our attention to this problem. This research was partly supported by the Pittsburgh Supercomputing Centre, the Alfred P. Sloan Foundation, NASA, NSF, and the Presidential Faculty Fellows Program. L.H. is an Alfred P. Sloan Fellow and Presidential Faculty Fellow.

Intercalation of ammonia into K_3C_{60}

M. J. Rosseinsky*, D. W. Murphy†, R. M. Fleming† & O. Zhou†

* Inorganic Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QR, UK

† AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill, New Jersey 07974, USA

THE superconducting transition temperatures (T_c s) in the alkali-metal-doped C_{60} superconductors (A_3C_{60}) seem to be determined by the size of the face-centred cubic (f.c.c.) unit cell, which in turn is dependent largely on the size of the metal cations occupying the interstitial sites¹. We have shown previously² that intercalation of ammonia into Na_2CsC_{60} expands the unit cell and thereby increases T_c by ~ 20 K. This has prompted us to explore the generality of this effect by attempting to expand the lattice of A_3C_{60} superconductors containing larger cations. Here we show that intercalation of ammonia into K_3C_{60} (which has a T_c of ~ 19 K) yields $(NH_3)K_3C_{60}$, which has a structure that can be described as an orthorhombic distortion of the f.c.c. structure with one K and one NH_3 per octahedral site. Despite strong evidence that the extent of charge transfer from K to C_{60} is unchanged by ammonia intercalation, the compound is not superconducting. We believe that the absence of superconductivity is a consequence of electron localization resulting from the structural distortion. Thus the change in symmetry of the unit cell, as well as the cell size, may be an important factor in attempts to create C_{60} -based superconductors with high T_c s.

Single phase K_3C_{60} (30 mg, with a superconducting volume fraction of 30%, as measured by d.c. magnetization) was placed in a 9-mm Pyrex tube, evacuated to 10^{-5} torr and exposed to NH_3 (at a pressure of 0.6 bar), distilled from liquid ammonia condensed on to sodium, using a vacuum manifold. After 20 min, the Pyrex ampoule was sealed under the ammonia pressure and then heated at 100°C for 10 days. The superconducting volume fraction decreased continuously to less than 0.1%. The tube was then opened in a dry box and the reacted sample sealed in an X-ray capillary under helium. Separate measurements of pressure changes on ammonia uptake and release on heating show the intercalation to be reversible and to correspond to the composition $(NH_3)_{0.95 \pm 0.05}K_3C_{60}$.

The X-ray diffraction pattern of the ammoniated K_3C_{60} (Fig. 1) revealed substantial changes when compared with the starting material in both the positions and intensities of the major reflections. Auto-indexing using the TREOR program (P.-E. Werner, L. Eriksson and M. Westdahl) gave a tetragonal unit cell of dimensions $a = 10.5 \text{ \AA} \times c = 13.7 \text{ \AA}$, and systematic absences consistent with body centring. The structural chemistry of A_3C_{60} can be unified using a body-centred tetragonal (b.c.t.) cell with $c/a = \sqrt{2}$ to represent the f.c.c. structure³. Intercalation of ammonia thus produces a slight distortion of a structure that remains predominantly f.c.c., with a 6% volume expansion. We therefore used an initial model with the potassium cations in the tetrahedral and octahedral sites of the f.c.c. structure, transformed into the b.c.t. cell.

The $I4/mmm$ space group was consistent with the systematic absences and was adopted for the initial refinements. The C_{60} must be disordered if the crystal system is tetragonal as the molecule has no four-fold axis. The presence of mirror planes with normals both along the cell vectors and also in the $\langle 110 \rangle$

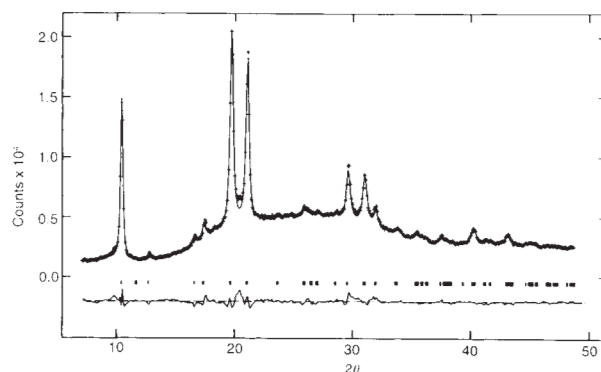


FIG. 1 Powder X-ray diffraction pattern of $(NH_3)K_3C_{60}$ in a sealed 0.5-mm glass capillary, collected with $Cu K\alpha$ radiation produced by a rotating anode source with a pyrolytic graphite focussing monochromator and flat pyrolytic graphite analyser. The crosses are experimental points and the solid line is a Rietveld fit to the model described in the text, with the difference shown on the same scale beneath the observed and calculated plots. Ticks mark the positions of the allowed Bragg reflections.

TABLE 1 Refined atomic coordinates for $(\text{NH}_3)_4\text{K}_3\text{C}_{60}$

Atom	Site	Symmetry	x	y	z	U (Å ²)	Occupancy
C(1)	16m	m..	0	-0.0468(2)	0.256(1)	0.020(9)	1.00
C(2)	16o	..m	0.0466(3)	-0.233(1)	0.0000	0.020(9)	1.00
C(3)	16n	.m.	0.2340(9)	0	0.0514(2)	0.020(9)	1.00
C(4)	32p	1	0.0784(2)	-0.0949(3)	0.2215(7)	0.020(9)	1.00
C(5)	32p	1	0.0951(3)	-0.2023(6)	0.0859(3)	0.020(9)	1.00
C(6)	32p	1	0.2024(7)	-0.0785(3)	0.1040(3)	0.020(9)	1.00
C(7)	32p	1	0.0482(2)	-0.1724(6)	0.1671(6)	0.020(9)	1.00
C(8)	32p	1	0.1741(7)	-0.1542(6)	0.0533(2)	0.020(9)	1.00
C(9)	32p	1	0.1527(6)	-0.0481(2)	0.1884(7)	0.020(9)	1.00
K(1)	8f	222	1/4	1/4	1/4	0.078(9)	1.00
K(2)	16o	..m	0.571(3)	0.937(1)	0	0.05(2)	0.25
N(1)	16o	..m	0.436(1)	0.052(2)	0	0.05(2)	0.25

$a = 14.971(7)$ Å, $b = 14.895(7)$ Å, $c = 13.687(4)$ Å. $R_w = 4.4\%$, $R_p = 3.1\%$, $R_i = 6.9\%$, $\chi^2 = 4.5$. Peak shape function: pseudo-Voigt. Background function: 8 term Chebyshev polynomial. Refinement was carried out using the GSAS Rietveld software (A. C. Larson and R. B. von Dreele, General Structural Analysis System, Los Alamos National Laboratory, 1985–1990).

directions means that there are two possible choices for the disordered orientations. Several intensities are not correctly fitted using a spherical shell form factor for C_{60} . The C_{60} orientation with C_2 along $\langle 100 \rangle$, found for K_4C_{60} (ref. 4) was unsatisfactory ($R_w = 20\%$) and we therefore considered the alternative merohedrally disordered model. In this, the C_{60} molecules are oriented with their C_2 axes parallel to the $\langle 110 \rangle$ directions, and the tetrahedral potassium is coordinated solely by six-membered rings. Refinement based on this model converged to $R_w = 9\%$. We then disordered the potassium onto the 8(j) $x, 1/2, 0$ position and varied both its position and occupancy. R_w decreased to 4.8% with an occupancy of 1.4 potassium cations on the octahedral site. This corresponds to precisely the same electron density that would arise from one potassium cation and one nitrogen (from an ammonia molecule) occupying the octahedral site in a disordered fashion.

This simple model gives rise to a square of scattering density whose vertices point towards the midpoints of the edges of the basal plane of the compressed octahedral site (Fig. 2, bottom). The diagonal of this square is 2.8 Å, corresponding to a physically reasonable estimate of the $\text{K} \cdots \text{N}$ distance (by comparison with the graphite intercalate $\text{K}(\text{NH}_3)_{4.33}\text{C}_{24}$ (ref. 5)). We conclude that one $\text{K}(\text{NH}_3)$ group occupies each octahedral site in the ammoniated compound, with disorder over the four possible orientations (the diagonally opposed corners of the square of scattering density).

The interplay of the disorder of the C_{60} molecules and of the $\text{K}(\text{NH}_3)$ group in the basal plane makes the identification of a physically acceptable local structure challenging. The 6:6 bond that lies in the basal plane may be oriented either parallel or perpendicular to that plane. $\text{N} \cdots \text{C}$ contacts of shorter than 2.7 Å result unless the in-plane 6:6 bond C_{60} orientation is forced not to be adjacent to an NH_3 molecule. This complication led us to search for more physically acceptable models to describe the bulk structure and allow a more satisfactory local description of the disordered $\text{K}(\text{NH}_3)$ complex.

A statistically equivalent refinement in $Fm\bar{3}m$ leads to more reasonable interatomic distances. In this model, alignment of the two-fold axes of the C_{60} molecule along the cell vectors produces just one molecular orientation in an orthorhombic cell derived by 45 degree rotation and doubling in volume of the tetragonal cell (that is, simply an orthorhombic distortion of the $Fm\bar{3}m$ parent structure⁶ with ordered C_{60} orientations). Refinement of the orthorhombic cell parameters converged smoothly to values which differed by over ten times their estimated standard deviations. The $\text{K}(\text{NH}_3)$ complex was initially modelled using a variable occupancy of N or K only, on the 16(o) $x, y, 0$ position, as for the tetragonal model. Refinement proceeded smoothly to $R_w = 4.7\%$. Most significantly, however, the $\text{N} \cdots \text{C}$ and $\text{K} \cdots \text{C}$ distances were all now greater than 3.0 Å, that is, the orthorhom-

bic distortion modifies the basal plane of the octahedral site to give physically reasonable $\text{C} \cdots \text{N}$ and $\text{C} \cdots \text{K}$ distances. Indeed, using soft bond length constraints only on the reasonably well defined closest contact $\text{C} \cdots \text{N}$ (3.2 Å) and $\text{K} \cdots \text{C}$ (3.06 Å in KC_8 (ref. 7)) distances, a model with quarter occupied, distinct K and N positions can be refined to give a $\text{K} \cdots \text{N}$ distance of 2.68(8) Å. Final refinements involved adjusting the radius of the sphere on which the carbon positions lie, subject to constraints of 1.39 ± 0.03 Å and 1.45 ± 0.03 Å on all the independent 6:5 and 6:6 bonds and $120 \pm 0.1^\circ$ and $108 \pm 0.1^\circ$ on the bond angles. (Fig. 1 and Table 1).

The reversibility of ammonia uptake indicates that amide (NH_2^-) formation does not occur and the C_{60}^- anions carry a 3- charge in the ammonia intercalate as well as superconducting K_3C_{60} . With powder X-ray diffraction, it is difficult to address the question of the hydrogen positions directly in this case due to the weak scattering power of H combined with the disorder. Calculations of the total van der Waals and electrostatic energy as a function of rotation angle (increased in five degree steps) about the C_3 axis of NH_3 (which was oriented along the N-K vector) showed that the most favourable orientation had a shortest $\text{C} \cdots \text{H}$ distance of 2.56 Å. This is similar to the shortest contact (2.58 Å) found for TDAE-C_{60} (ref. 8) and shorter than common $\text{C} \cdots \text{H}$ non-bonded distances.

We emphasize that the scattering density can be accounted for satisfactorily using an unconstrained model with one disordered species of variable occupancy on the octahedral site in $I4/mmm$. The $Fm\bar{3}m$ refinement with bond length constraints, however, allows a more satisfying description of the bonding on the disordered octahedral site (Fig. 2, top). In K_3C_{60} , the octahedral cation is equidistant from the 6:6 ring fusions of the six neighbouring C_{60} molecules, which adopt one of two orientations (12×3.69 Å $\text{C} \cdots \text{K}$ distances) (ref. 6). The observed scattering density here indicates that the potassium is displaced towards the midpoint of the edges of the square basal plane of the compressed octahedral site (note that the magnitudes of the lattice parameters are consistent with our structural model), with the coordinating ammonia 2.7 Å away, displaced from the centre of the octahedral site towards the opposite edge (Fig. 2). Four C_{60}^{3-} anions (two along the c -axis and two in the basal plane) and the ammonia form a distorted trigonal bipyramidal environment around the potassium cation. There are two 3.5 Å contacts to one of the 6:6 carbons (C(1)) from each of the neighbouring anions along the c axis. In the basal plane, the potassium is now closest to two of the four equatorial C_{60} s. In $Fm\bar{3}m$, the orientations of the 6:6 bond in these two molecules with respect to the basal plane differ: one anion has the 6:6 bond lying in the plane (C(2)) and forms a 3.05 Å interaction with the cation (shorter than the 3.3 Å found on the tetrahedral site). The other C_{60} closest to K has its 6:6 bond oriented perpendicular to the

basal plane, producing two equidistant $K \cdots C(3)$ 3.22 Å interactions. The nitrogen is displaced towards the two remaining C_{60} molecules in the basal plane of a given octahedral site. The preference we have for the *Fmm* model is purely on the basis of bond length: the $C \cdots N$ distances in the tetragonal model are 0.6 Å too short to be chemically reasonable.

The geometry and stoichiometry of NH_3 co-ordination to K^+ on the octahedral site in the present case differs dramatically from that to Na^+ in superconducting $(NH_3)_4Na_2CsC_{60}$. Chemically, the large potassium cation would be expected to coordinate less strongly with NH_3 than sodium. Possible explanations for the resulting differences in physical behaviour have wide implications both for the electronic structure of these systems, and the prospects for significantly enhanced T_c s. We note that all previous fulleride superconductors are cubic, and the reduction in symmetry and dimensionality (which favours electronically-driven charge or spin density wave transitions) may be responsible for the absence of superconductivity here. This would imply

that simply enhancing the cell volume without maintaining cubic symmetry will be insufficient to give higher T_c s. In fact, the unit cell volume of $(NH_3)_3K_3C_{60}$ is almost identical to that of Rb_2CsC_{60} (ref. 1). Alternatively, the low symmetry distortion could be driven by electron localization produced by the expansion. Determination of physical properties is required to investigate whether sufficient band narrowing occurs to make $(NH_3)_3K_3C_{60}$ a Mott-Hubbard insulator, in which the t_{1u} electrons are localized by correlation. □

Received 11 March 1993; accepted 7 June 1993.

1. Fleming, R. M. et al. *Nature* **352**, 787–788 (1991).
2. Zhou, O. et al. *Nature* **362**, 433–435 (1993).
3. Murphy, D. W. et al. *J. Phys. Chem. Solids* **53**, 1321–1332 (1992).
4. Zhou, O. & Cox, D. E. *J. Phys. Chem. Solids* **53**, 1373–1390 (1992).
5. Solin, S. A. *J. Phys., Paris IV* **C8**, 311–324 (1991).
6. Stephens, P. W. et al. *Nature* **351**, 632–634 (1991).
7. Rudorff, W. *Adv. inorg. Chem. Radiochem.* **1**, 223–266 (1959).
8. Stephens, P. W. et al. *Nature* **355**, 331–332 (1992).

ACKNOWLEDGEMENTS. We thank A. P. Ramirez, R. Tycko and R. C. Haddon for discussions and experimental assistance.

Femtosecond dynamics of dissociation and recombination in solvent cages

Qianli Liu, Juen-Kai Wang & Ahmed H. Zewail

Arthur Amos Noyes Laboratory of Chemical Physics,
California Institute of Technology, Pasadena, California 91125, USA

FOR chemical reactions in solution, the solvent exerts an important influence on the elementary processes of bond making and breaking. The solvent may, for example, enhance bond formation by trapping reactive species in a 'solvent cage' on the reaction timescale¹, or it may act as a 'chaperone' that stabilizes energetic species². Ultrafast reaction dynamics in solvent shells can be probed using laser spectroscopic techniques developed to resolve atomic motion on the femtosecond (fs) timescale³. Here we report on a study of the femtosecond dynamics of the dissociation of neutral iodine molecules encaged in clusters of around 40–150 argon atoms, which form a solvent shell^{4–6}. We find that, when dissociation occurs from the A-type excited electronic state of I_2 , the iodine atoms exhibit coherent motion on a sub-picosecond ($<10^{-12}$ s) timescale, rebounding from the 'frozen' solvent cage and recombining. The 'hot' I_2 molecule is then cooled over by collisions with the argon atoms. We provide support for these interpretations using molecular-dynamics simulations. Dissociation from the B state, meanwhile, involves slower bond-breaking and slower recombination of the fragments—there is no coherent 'rebound' from the solvent cage. The dissociation pathway therefore depends critically on the timescale of bond breaking relative to that of solvent rearrangement.

Direct real-time studies of reactions in solvent clusters on the picosecond^{7–9} and femtosecond^{10–14} timescales, along with molecular-dynamics simulations^{15–17}, have revealed the dependence of reaction dynamics on the number of solvent atoms or molecules in the cluster. Iodine has been studied in molecular-beam experiments involving solvent-like clusters ranging from many atoms^{4,18} to a single atom^{19–21} in size; these studies provide insights into such processes occurring in bulk condensed phases^{22–25}. Building on our previous work on the ultrafast dynamics of isolated iodine molecules^{26,27}, here we adopt this molecular-beam approach to monitor the dissociation and atom recombination dynamics in a solvent cage on the femtosecond timescale.

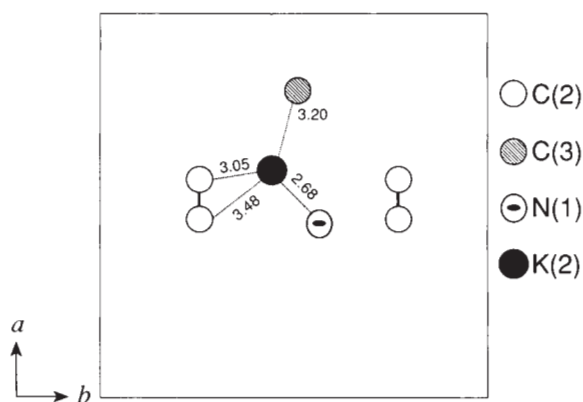
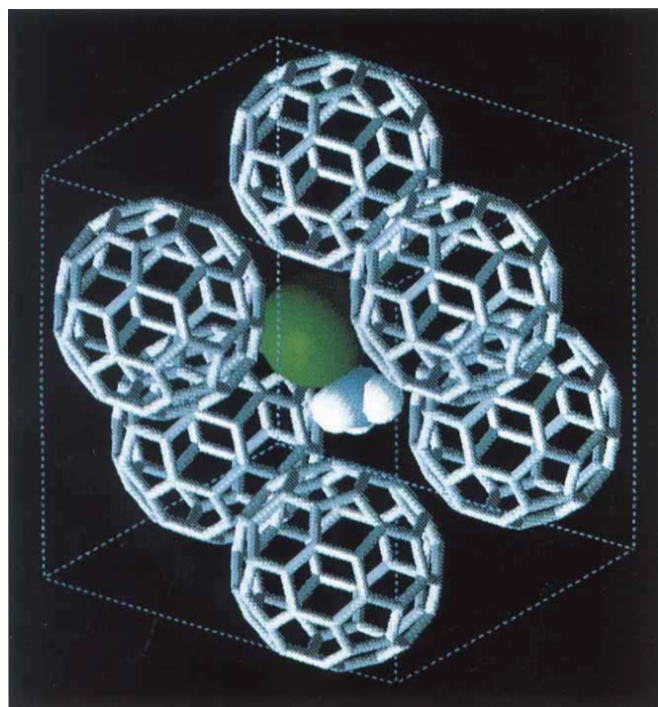


FIG. 2 Top, the octahedral site of $(NH_3)_3K_3C_{60}$: C, K, N and H are represented in grey, green, blue and white respectively. Bottom, the coordination to potassium in the basal plane. All the atoms shown are at $z = 0$, except C(3) which represents two carbon atoms in a 6:6 bond at $z = \pm 0.051$ projected into the basal plane. The centre of the inset is at $(1/2, 0, 0)$. Distances are marked in Å.

We encaged neutral iodine molecules in argon clusters in a molecular beam. Dissociation was initiated by a femtosecond (fs) laser pulse which prepares a wave packet either on the A state above its dissociation limit (to $I+I$) or on the B state at different energies below or above dissociation (to $I+I^*$). The timescales for bond breaking in these two states are very different. As shown in Fig. 1, on the A potential, the dissociation in the gas phase is direct and results in a large available translational energy for the I atoms. On the B potential, indirect dissociation (predissociation) occurs, involving crossing from a bound to a repulsive surface. In a solvent cage, there is a 'solvent barrier' to dissociation (Fig. 1).

A second fs pulse is used to probe^{26,27} the motion of the wave packet by detecting¹⁰ the laser-induced fluorescence of I_2 in the Ar clusters at its characteristic emission¹⁸. The probing windows on the potentials correspond to known transitions to ion-pair states¹⁸. From electron diffraction, the average size of the clusters formed is estimated to range from 40 to 150 argon atoms, with an average temperature of 30 K or less²⁸. To compare the dynamics in the clusters to those of bare iodine^{26,27}, argon gas was replaced by helium which is known to form no such clusters under these expansion conditions.

Figure 2 shows experimental transients obtained when iodine molecules were initially excited at 614 nm, typical of A-state behaviour. The B-state contribution in the cluster at this wavelength is insignificant, as we observe similar behaviour from 590 to 700 nm. In Fig. 2a, the first peak rise characterizes the preparation of the wave packet. The following decay reflects the fact that the wave packet is moving into a region of large separations where iodine can no longer absorb the probe pulse. In about 250 fs, the signal drops to almost zero. A fast and prompt recovery was then observed in another ~ 300 fs, which indicates that the wave packet coherently moves back to optical regions of the probe. This recovery is comparable to the coherent

motion of isolated I_2 in a bound potential^{26,27}. Here it is the argon cage that provides the 'outer' boundary of the potential well. At longer times, the signal starts to increase almost linearly and much more slowly, as shown in Fig. 2b. The molecular dynamics (MD) simulations discussed below show that this increase is due to recombination of the two iodine atoms ($I+I$) on the ground state potential. A steady-state value is approached for this recombination process in about 30 ps. The inset in Fig. 2a gives the corresponding transients obtained for bare iodine in helium expansions ($\lambda_{\text{pump}} = 614$ and 640 nm) and shows the direct dissociation at these excitations²⁹.

Excitation into the B state (Fig. 3) leads to totally different behaviour; the initial fs dynamics (discussed above) is replaced by a picosecond decay (~ 15 ps). Following this decay, the signal rises again in ~ 30 ps, as given in Fig. 3 for $\lambda_{\text{pump}} = 520$ nm. In the inset of Fig. 3 (top), a transient obtained by the same pump and probe (310 nm) is shown for the helium expansion. This transient shows the ~ 1 ps modulation which reflects the vibrational period of bare iodine molecules on the B state at this energy^{26,27}. Above the B-state dissociation into $I+I^*$, we also observed caging (Fig. 3, bottom).

These results demonstrate that in argon clusters, the motion of the wave packet is dramatically different from that of bare iodine and clearly reflects the initial coherence and the caging dynamics. To help understand details of the motion in the solvent cages, we have performed MD simulations. The clusters were assumed to have 17 or 44 argon atoms, which form one or two argon shells, with the iodine molecule at the centre¹⁰. A simulated transient was obtained by accumulating the time when the trajectories of the iodine molecules were within the probe windows (these results will be detailed elsewhere). Figure 4 shows the simulated signals for the A-state and B-state dynamics; these simulations reproduce the essential features of the experimental results.

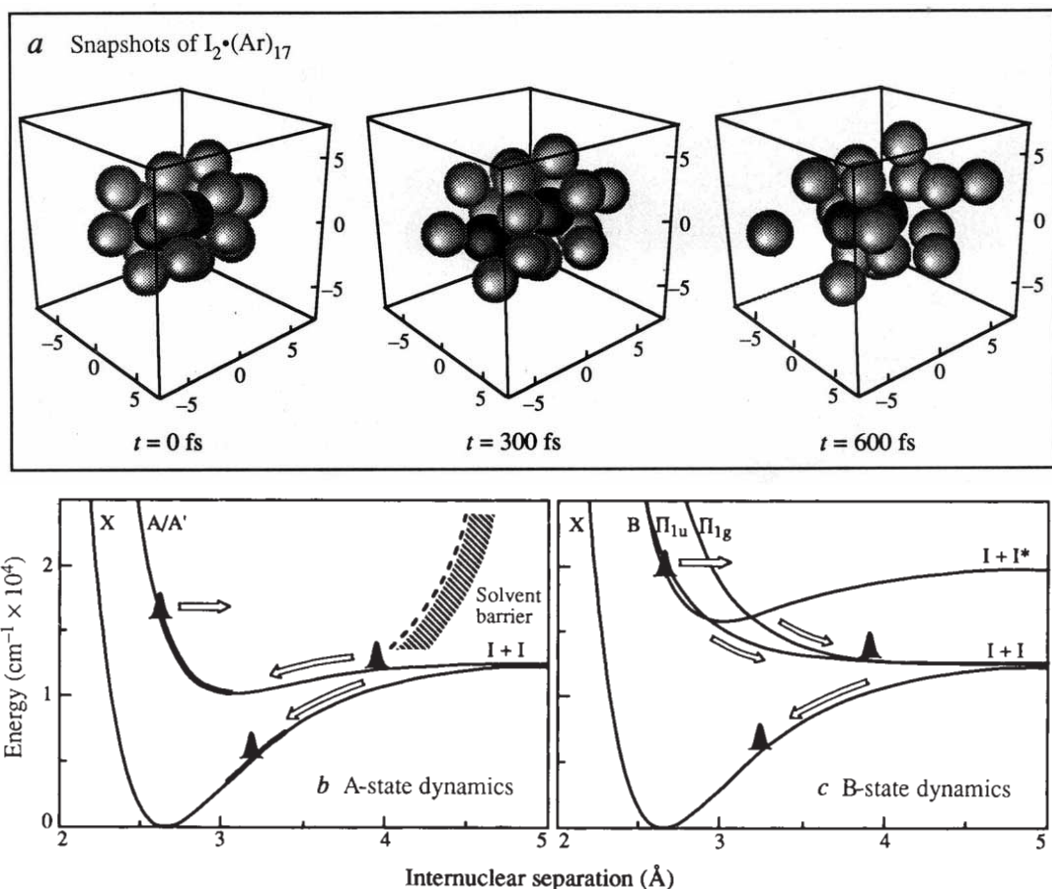


FIG. 1 a, Snapshots of the structures of iodine (dark grey) in argon (light grey) solvent cages (17 atoms) at the first 600 fs. The atomic sizes are chosen to enhance the structural changes visually. Iodine was initially prepared on the A state above its dissociation limit (614 nm). The separation of $I+I$ at ~ 300 fs and direct recombination at ~ 600 fs are two snapshots of the motion on the potential given below. b, c, The potential energy surfaces of the iodine molecule, together with snapshots of a wave packet motion on the A state (b) or the B state (c). The dashed line in conjunction with the dashed area indicates the equivalent solvent barrier caused by the repulsion with Ar atoms (see text), and it is similar for both b and c.

We can now develop a microscopic picture of the motion in the solvent cages. From the MD simulations of the A-state type behaviour (Fig. 4a), we can see that the bouncing-back of the wave packet from the solvent barrier (along the A/A' potential) results in a prompt recombination within 0.6 ps. On this fs timescale, the solvent is still 'frozen', and from the observed time delay and MD simulations (see Fig. 4a) the effective barrier is at ~ 4.4 Å. At longer times, the simulations show that for recombination on the ground X state (and not the A/A' state) the signal rises almost linearly. The ps rise reflects the fact that the two iodine atoms are vibrationally hot when they first recombine at long internuclear distances (4–5 Å; Fig. 1). Energy transfer to argon through periodic collisions (cooling) allows the iodine to relax to shorter internuclear distances (with better Franck-Condon factors for absorption to the ion-pair states). Equilibrium is approached after ~ 30 ps. The linear increase reflects this loss of energy to the solvent in steps. The time evolution of the intrinsic $\langle I-I \rangle$ distance, also shown in Fig. 4, clearly indicates coherent motion on the timescale of observation.

For the B state (Fig. 4b), the simulations do not show the initial femtosecond fall and rise—both decay of the B state (via predissociation) and recombination on the ground state occur on the ps time scale. Because in this case the timescale for bond

breakage is several picoseconds, the recovery reflects an essentially equilibrated distribution (near-exponential growth), as the solvent cluster has enough time to absorb the energy of the dissociating iodine atoms and soften its structure. Once dissociated, the two iodine atoms separate to long internuclear distances, allowing solvent atoms to intervene between them. Above the B-state dissociation limit, we also observe a caging effect, in addition to a quasi-bound B-state contribution (Fig. 3). A full account of these results will be presented elsewhere.

These studies point to several conclusions. First, the timescale for bond-breaking is critical to the subsequent bond-making (caging) dynamics. Second, the caging phenomenon involves initial coherent femtosecond motion followed by cooling through energy transfer to the solvent⁴. Third, the solvent temperature is an integral part of the recombination² dynamics. In liquids^{23,30}, the motion of the solvent at higher temperature could mask the elementary coherent motion observed here. In a matrix at low temperatures, on the other hand, the rigidity should allow for the observation of the direct motion¹⁶, provided that the matrix phonons do not induce energy relaxation and dephasing of the coherent wave packet. Finally, the analogy with the previous study¹¹ on ionic clusters, $I_2^+ \cdot (CO_2)_n$, is significant. The coherent motion studied here for the dissociation (wave packet decay) and recombination of neutral $I_2 \cdot Ar_n$ reflects the short-

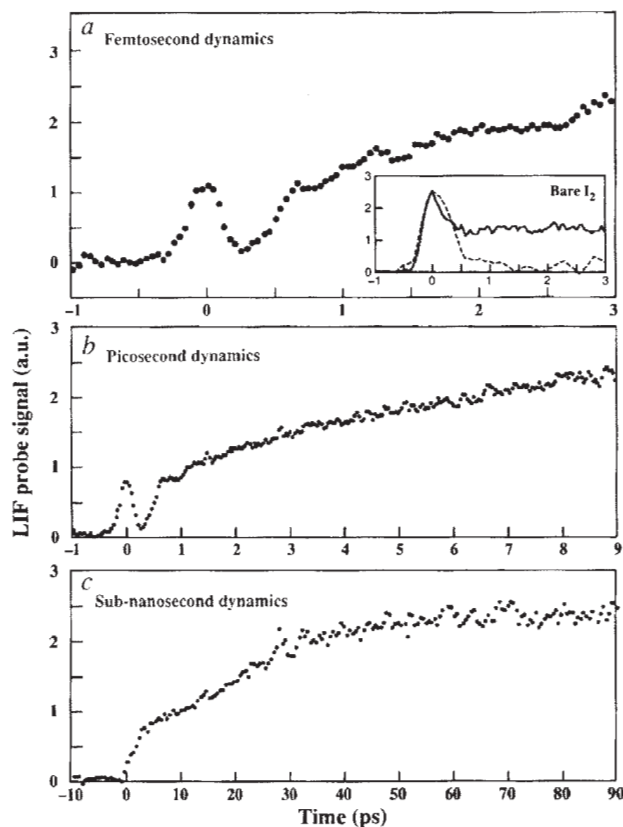


FIG. 2 Experimental femtosecond to sub-nanosecond transients of the iodine caged in Ar clusters, typical of A-type behaviour. The pump wavelength was tuned to 614 nm, and the probe wavelength was 310 nm. The inset in a shows the corresponding transients of bare iodine molecules (in helium expansions) using the same pump (614 nm) and probe (310 nm) laser pulses (solid line) and using a different pump wavelength (640 nm, dashed line). Note that at 614 or 640 nm the excitation is above the A-state dissociation to $I+I$ (at 614 nm, the constant background, at long times, is because of some bound state B excitation²⁹). b, The near-linear growth of the signal on the ps timescale. c, The signal approaches a levelling value at ~ 30 ps later (the LIF probe signal is in arbitrary units, a.u.).

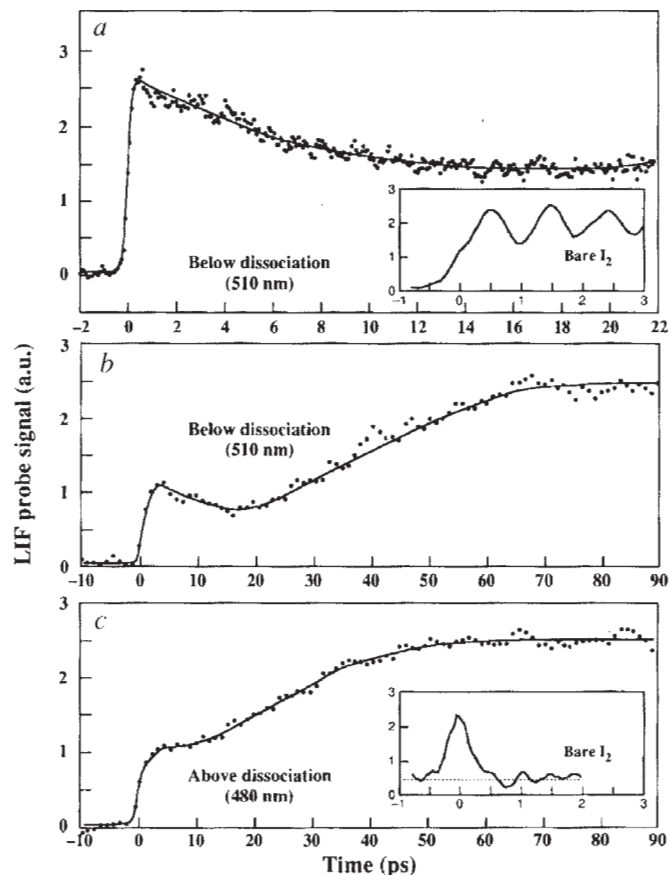


FIG. 3 Experimental transients obtained for the B state for pump wavelengths of 510 nm and 480 nm and a probe wavelength of 310 nm. The inset in a shows the corresponding transient for bare iodine, which displays the oscillatory motion for the bound state at 510 nm excitation. The two transients in a and b are for the same excitation (510 nm), but for different timescales. Here predissociation occurs by crossing to a repulsive potential. The transient in c is for excitation (480 nm) above the dissociation limit of the B state. The inset for this (c) transient shows, on a short timescale, the corresponding transient obtained for bare iodine in helium expansions.

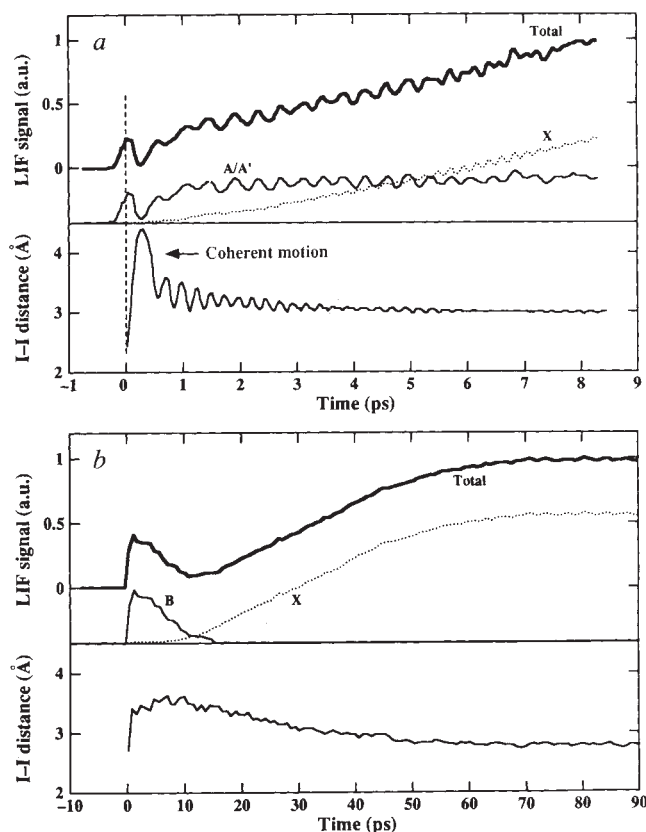


FIG. 4 Laser-induced fluorescence signals and internuclear separation calculated from molecular dynamics simulations. *a*, Simulated transients of $I_2 \cdot Ar_{44}$ when iodine is excited to the A state above its dissociation limit by a pump wavelength of 614 nm. The dotted line is the contribution from recombination on the ground state X. The thin solid line comes from recombination on the A/A state. The thick solid line is the sum (see text). *b*, Simulated transients with pump wavelength tuned to 510 nm which excites the B state. The thin solid line represents signal from B state, while the dotted line represents contributions from the X state. The thick solid line is the sum. For both A- and B-state dynamics, the intrinsic change of the I-I distance with time is shown. Note the dramatic change in the motion on A potential compared to the B potential, and that the coherent motion is evident only in the A-state dynamics. The simulations invoke classical mechanics and the persistence of some oscillations up to 9 ps might disappear in a quantum calculation²⁷.

range nature of the interaction with the solvent, while for ionic systems the long-range Coulomb interaction dominates. As with isolated chemical reactions³ such studies of coherent nuclear motions^{24,31,32} in complex systems of solvent cages allow for resolution of the dynamics on the timescale of atomic motion. □

Received 22 April; accepted 7 July 1993.

1. Frank, J. & Rabinovitch, E. *Trans. Faraday Soc.* **30**, 120–131 (1934).
2. Porter, G. & Smith, J. A. *Proceedings Royal Soc. A* **261**, 28–37 (1961).
3. Zewail, A. H. *Faraday Discuss. Chem. Soc.* **91**, 207–237 (1991).
4. Saenger, K. L., McClelland, G. M. & Herschbach, D. R. *J. phys. Chem.* **85**, 3333–3337 (1981).
5. Jortner, J., Scharf, D., Ben-Horin, N., Evan, U. & Landman, U. in *Proc. int. School of Physics* (ed. Scholes, G.) 43–98 (North-Holland, Amsterdam, 1990).
6. Berry, R. S. in *Proc. Int. School of Physics* (ed. Scholes, G.) 3–22 (North-Holland, Amsterdam, 1990).
7. Papanikolas, J. M. et al. *J. phys. Chem.* **95**, 8028–8040 (1991).
8. Breen, J. J. et al. *J. chem. Phys.* **92**, 805–807 (1990).
9. Gutmann, M., Willberg, D. M. & Zewail, A. H. *J. chem. Phys.* **97**, 8048–8059 (1992).
10. Potter, E. D., Liu, Q. & Zewail, A. H. *Chem. Phys. Lett.* **200**, 605–614 (1992).
11. Papanikolas, J. M. et al. *J. chem. Phys.* **97**, 7002–7005 (1992).
12. Baumert, T., Engel, V., Meier, C. & Gerber, G. *Chem. Phys. Lett.* **200**, 488–494 (1992).

13. Wei, S., Purnell, J., Buzza, S. A., Stanley, R. J. & Castleman, A. W. Jr *J. chem. Phys.* **97**, 9480–9482 (1992).
14. Schreiber, E., Kühling, H., Kobs, K., Rutz, S. & Wöste, L. *Ber. Bunsenges. phys. Chem.* **96**, 1301–1305 (1992).
15. Amar, F. G. & Berne, B. J. *J. phys. Chem.* **88**, 6720–6727 (1984).
16. Alimi, R., Gerber, R. B., McCaffray, J. G., Huntz, H. & Schwentner, N. *Phys. Rev. Lett.* **69**, 856–859 (1992).
17. Stace, A. J. *J. Chem. Soc. Farad. Trans. II* **77**, 2105–2110 (1981).
18. Fei, S., Zheng, X., Heaven, M. & Tellinghuisen, J. *J. chem. Phys.* **97**, 6057–6063 (1992).
19. Valentini, J. J. & Cross, J. B. *J. chem. Phys.* **77**, 572–573 (1982).
20. Beswick, J. A., Monot, R., Philippoz, J.-M. & van den Bergh, H. *J. chem. Phys.* **86**, 3965–3967 (1987).
21. Burke, M. L. & Klemperer, W. *J. chem. Phys.* **98**, 1797–1809 (1993).
22. Schroeder, J. & Troe, J. A. *Rev. phys. Chem.* **38**, 163–190 (1987).
23. Harris, A. L., Brown, J. K. & Harris, C. B. A. *Rev. phys. Chem.* **39**, 341–366 (1988).
24. Scherer, N. F., Ziegler, L. D. & Fleming, G. R. *J. chem. Phys.* **96**, 5544–5547 (1992).
25. Dardi, P. S. & Dahler, J. S. *J. chem. Phys.* **98**, 363–372 (1993).
26. Bowman, R. M., Dantus, M. & Zewail, A. H. *Chem. Phys. Lett.* **161**, 297–302 (1989).
27. Gruebele, M. & Zewail, A. H. *J. chem. Phys.* **98**, 883–902 (1993).
28. Farges, J., de Ferand, M. F., Raoult, B. & Torchet, G. *J. chem. Phys.* **84**, 3491–3501 (1986).
29. Tellinghuisen, J. *J. chem. Phys.* **58**, 2821–2834 (1973).
30. Nesbit, D. J. & Hynes, J. T. *J. chem. Phys.* **77**, 2130–2143 (1982).
31. Joly, A. G., Ruhman, S., Kohler, B. & Nelson, K. A. *Springer Ser. chem. Phys.* **48**, 506–510 (1988).
32. Benjamin, I., Banin, U. & Ruhman, S. *J. chem. Phys.* **98**, 8337–8340 (1993).

ACKNOWLEDGEMENTS. This work was supported by a grant by the National Science Foundation.

Template mineralization of self-assembled anisotropic lipid microstructures

Douglas D. Archibald & Stephen Mann*

School of Chemistry, University of Bath, Bath BA2 7AY, UK

THE high-fidelity replication and functional specificity of biomaterials and biopolymers is inspiring a renewal in the development of biomimetic strategies for materials synthesis^{1,2}. Approaches involving the templating or nucleation of inorganic materials by compressed monolayers of amphiphiles^{3–5}, the use of supramolecular lipid or protein cages in the preparation of nanoscale inorganic structures^{6–8} and the mineralization of bacterial fibres⁹ indicate the potential of controlled crystallization at inorganic–organic interfaces. Here we report the controlled formation of tubular inorganic–organic composites by using self-assembled lipid tubules as templates for the crystallization of inorganic oxides. We use microstructures formed by a sugar-based lipid galactocerebroside, doped with small amounts of an anionic sulphated derivative, to induce nucleation of magnetic and non-magnetic iron oxides. By varying the reaction conditions, we can create either tube-like or lamellar disk-like composites. Our results suggest that the variety of microstructures formed by chemically modified sugar-based lipids may provide a route to the production of mineral-containing fibres and other ceramic–organic composites.

To prepare oriented mineralized microstructural units that can be processed into higher-order assemblies, isotropic amphiphilic structures such as Langmuir monolayers (in two dimensions) and vesicles (in three) are of limited use—what one needs are anisotropic structures such as tubes or fibres. Self-assembled lipid tubules¹⁰ with large axial-ratios have been observed in systems containing semisynthetic chiral amphiphiles¹¹, amino-acid-based surfactants^{12,13}, sugar-based surfactants¹⁴ and biological galactocerebroside¹⁵. Recent studies have shown that tubules containing phosphatidylcholine headgroups could be metallized with electroless nickel¹⁶, coated in preformed colloidal silica¹⁷, or mineralized with a metal carbonate by chemical precipitation¹⁸. The latter process, although based on surface chemical interactions, was not highly specific to the organic sur-

* To whom correspondence should be addressed.

face. Here we describe the chemical tailoring of a newly discovered large axial-ratio lipid microstructure to produce nucleation sites for the synthesis of inorganic-organic composites of anisotropic (fibrous) morphology. Our approach is outlined in Fig. 1. Mixtures of galactocerebrosides such as α -hydroxy fatty acid galactocerebroside (HFA-Cer) and non-hydroxy fatty acid galactocerebroside (NFA-Cer) form a range of microstructures (I, II and III in Fig. 1) depending on lipid composition (Fig. 1b). The details of these microstructures and phase relationships are described in detail elsewhere¹⁹. These microstructures are activated for inorganic mineralization by the incorporation of low levels of anionic sulphated galactocerebroside (S-Cer, sulphatide) into the neutral lipid microstructures of HFA-Cer and NFA-Cer, without significant disruption of their morphologies¹⁹.

Lipid cylinders of variable composition were prepared by thermal-mechanical cycling of 1.5 mg ml⁻¹ suspensions of mixtures (% by weight) of bovine brain HFA-Cer (Sigma) and bovine brain S-Cer (Lipid Products) in ethylene glycol as

described^{15,19}. The cylinders were open-ended multilamellar tubules with a diameter of ~120 nm and a central lumen of 10–30 nm (Fig. 2a). Aliquots (40 cm³) of the lipid suspension in ethylene glycol were incubated with 4–100 cm³ of freshly prepared 10 mM FeCl₃ in 0.10 M HCl at room temperature, under air. The best conditions for specific mineralization of the tubules were found to be: 0.9 mM FeCl₃, 1 mM HCl and 0.06 mM of 90% HFA-Cer : 10% S-Cer. Mineralization resulted in the slow appearance of an orange colour, coagulation, and sedimentation of the coated lipid tubules. Significantly, controls without iron did not cause lipid coagulation or sedimentation, and microstructures not doped with S-Cer (that is 100% HFA-Cer) sedimented at a slower rate and did not develop the orange colour.

The mineralized tubules were covered in granular and sheet-like material (Fig. 2b) which gave peaks for oxygen, sulphur and iron by energy dispersive X-ray analysis. Selected-area electron diffraction patterns recorded on individual tubules indicated that the inorganic coating was poorly crystalline lepidocrocite (γ -FeOOH) (d_{hkl} : 0.331 nm (120), 0.254 nm (031), 0.242 nm (111)

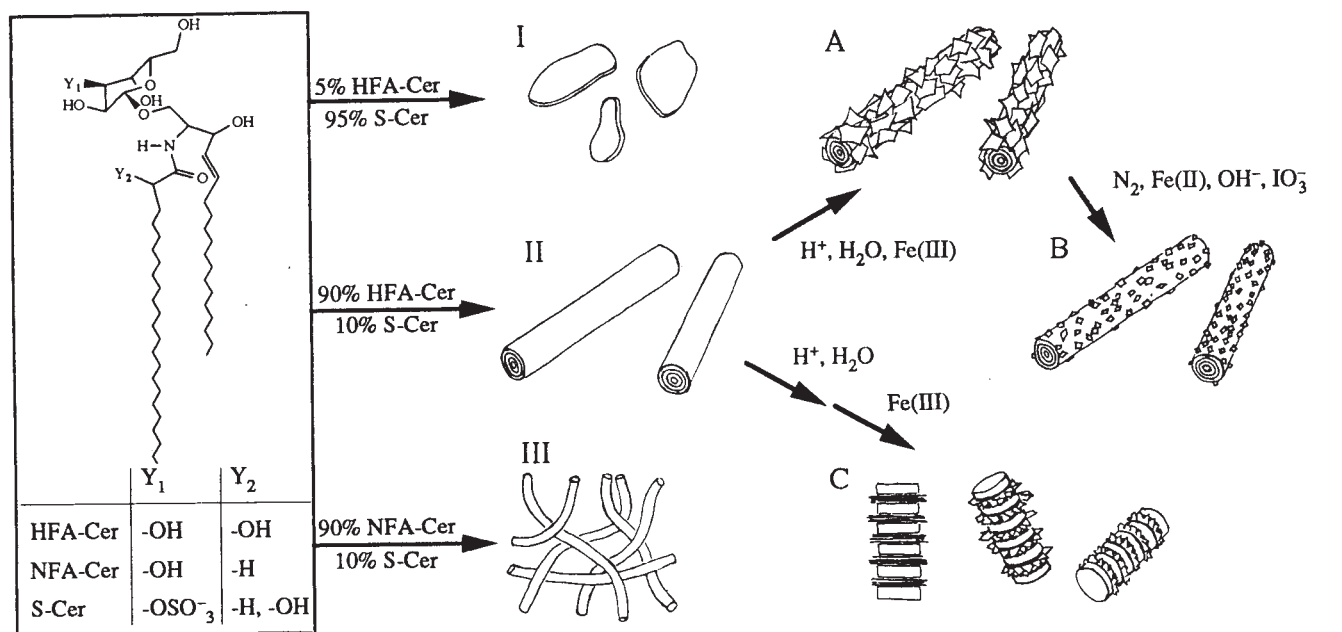
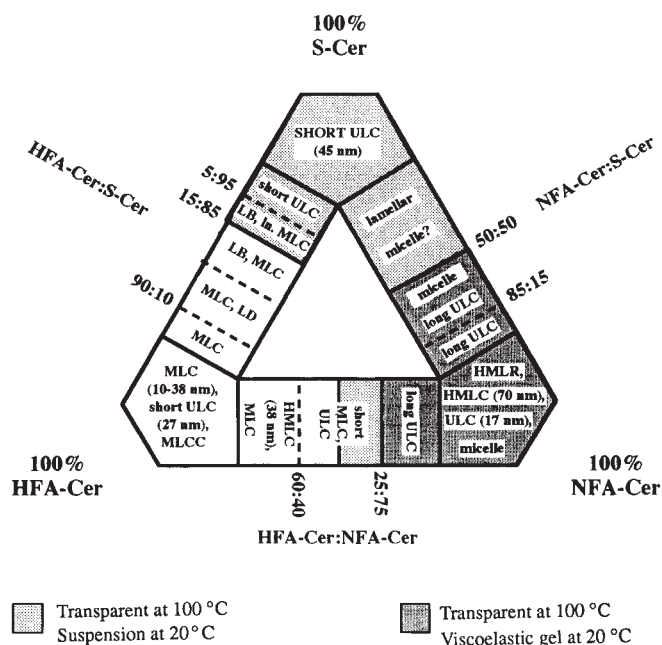


FIG. 1 a, Generalized scheme for the introduction of shape anisotropy in biomimetic inorganic-organic composites. Galactocerebroside mixtures can produce a variety of microstructures in ethylene glycol suspension: I, lamellar disks; II, multilamellar tubules; III, Viscoelastic gel of fibrous unilamellar tubules. In synthetic route A, a cylindrical coating of γ -FeOOH (lepidocrocite) is generated by incubation of microstructures II in a solution of acidic Fe(III). Synthesis B illustrates that the γ -FeOOH on these tubules can be subsequently converted *in situ* to a ferrimagnetic inverse spinel material (magnetite or maghemite) by room temperature anaerobic reaction with stoichiometric amounts of Fe(II)²¹. In both routes, hydration of the bilayer tubule structures is retarded by binding of Fe(III) and iron oxides, but if route C is followed, S-Cer-rich lamellar disks separate from the parent microstructures by suspension in acidic water before addition of Fe(III). Subsequent binding of Fe(III) by these disks results in growth of oriented thin plates of γ -FeOOH crystals intercalated between the lipid lamellae. b, Diagrammatic summary of phase behaviour exhibited in mixtures of the lipids α -hydroxy fatty acid galactocerebroside (HFA-Cer), non-hydroxy fatty acid galactocerebroside (NFA-Cer) and anionic sulphated galactocerebroside (S-Cer). Each corner of the diagram describes the microstructures produced by the pure lipid prepared by slow cooling of glycol suspensions at 100 °C. LB=lamellar blebs; LD=lamellar disks; ULC=unilamellar cylinders; MLC=multilamellar cylinders; HMLC=helically patterned multilamellar cylinders; HMLR=helical multilamellar ribbons; MLCC=multilamellar cochleate cylinders. Numbers in parentheses are typical lumen diameters.



and 0.121 nm (222)) with a small fraction of ferrihydrite ($\text{Fe}_2\text{O}_3 \cdot n\text{H}_2\text{O}$). The sheet-like material was identified as γ - FeOOH by high resolution lattice images which clearly showed a 0.66 nm spacing (d_{hkl} (020)). FT-infrared spectroscopy of the lipid-mineral composite showed the 740 cm^{-1} characteristic band of lepidocrocite (data not shown). Infrared spectroscopy also gave results which implied that the hydrocarbon and amide functionalities were largely unchanged within the molecular structure associated with the mineralized tubules, whereas the galactosyl interface was partially modified due to the presence of iron oxyhydroxide and/or water of hydration.

Conversion of lepidocrocite coated tubules *in situ* was carried out under nitrogen, at room temperature and with minimal stirring. A small volume of deaerated $(\text{NH}_4)_2\text{SO}_4 \cdot \text{Fe}(\text{II})\text{SO}_4 \cdot 6\text{H}_2\text{O}$ solution was added to a suspension of the $\text{Fe}(\text{III})$ mineral-coated tubules to give a 1:1 molar ratio of $\text{Fe}(\text{II})$: $\text{Fe}(\text{III})$. The phase transformation reaction was then initiated by addition of 100 mM NaOH to raise the pH to 8.5. In some experiments, iodate (IO_3^-) was also added as an oxidant to produce tubules extensively mineralized with crystals of poorly-ordered goethite (α - FeOOH) (Fig. 3a). In other experiments, the reaction mixture was maintained at pH 8–9 by additional aliquots of NaOH and left for 4 days. A black tubule-containing precipitate, which was responsive to the field of a bar magnet, was isolated. Electron microscopy images showed that the lipid tubules were encrusted with clusters of fine-grained inorganic material (Fig. 3b) which was identified by electron diffraction (d_{hkl} : 0.479 nm (111), 0.300 nm (220)) as the ferrimagnetic inverse spinel phase magnetite (Fe_3O_4). The electron diffraction data were not sufficient to distinguish this mineral from the oxidized ferrimagnetic spinel, maghemite (γ - Fe_2O_3).

Different results were obtained if a S-Cer-doped lipid tubule

suspension in ethylene glycol was incubated for 10 min in acid solution before addition of $\text{Fe}(\text{III})$ (route C, Fig. 1). This produced a S-Cer-rich phase which was observed by TEM to be formed by partial phase separation from the parent tubules. Subsequent $\text{Fe}(\text{III})$ addition caused aggregation of the segregated anionic lipid into stacks of lamellae, interspaced with nanometre-thin layers of poorly-crystalline iron oxyhydroxide. Presumably, $\text{Fe}(\text{III})$ binds preferentially to the surface of the disks because of their high negative charge, and the concomitant neutralization of surface charge gives rise to stack formation. The initial mineral was transformed over a period of weeks into sheet-like crystals of lepidocrocite which were aligned by confinement between the stacked lipid lamellae (Fig. 3c). The stacked disk nanocomposite was also produced directly in pure S-Cer controls but the product was less well-defined (data not shown). As the disks are ~ 6.5 nm thick, the mineralized system is a nanocomposite of inorganic and organic layers. Although more structural work is required to determine the details of this association, it is possible that organized stacks of inorganic-organic materials could be of potential significance in the catalytic, photochemical and electronic application of finely divided solids.

Our results indicate that high axial ratio inorganic-organic composites can be routinely produced by template mineralization of chemically-modified galactocerebroside tubules. Incorporation

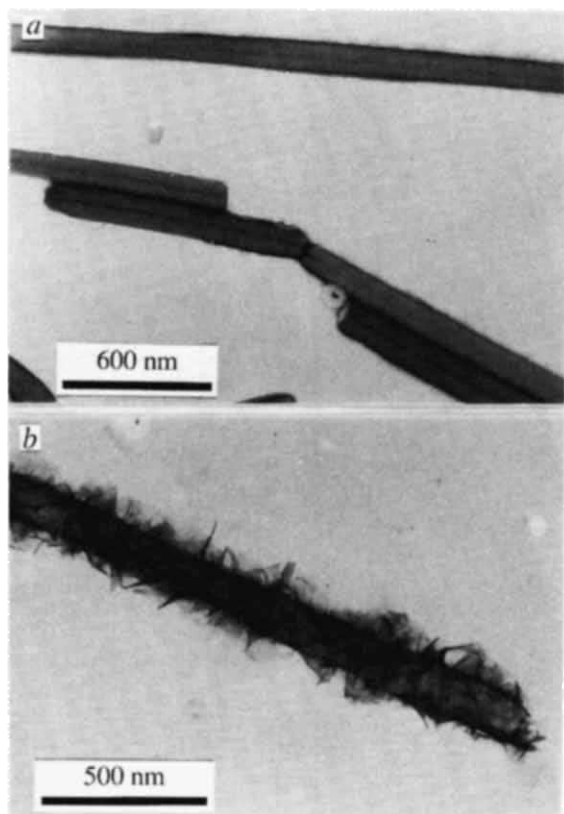


FIG. 2 Transmission electron micrographs of lipid tubules and mineralized products. a, Lipid tubules formed from a mixture of 90% HFA-Cer and 10% S-Cer prior to mineralization. b, Mineralization with lepidocrocite by synthetic route A (Fig. 1a).

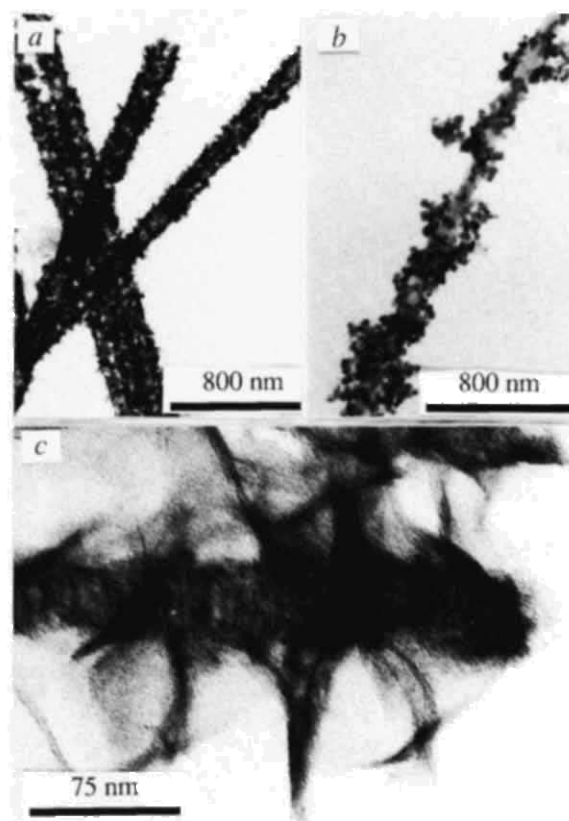


FIG. 3 Transmission electron micrographs of lipid-mineral composites after *in situ* conversion reactions. a, α - FeOOH encrusted tubules (90% HFA-Cer, 10% S-Cer) formed by secondary oxidation of $\text{Fe}(\text{II})$ on the surface of γ - FeOOH coated fibres. b, Magnetic tubules formed by reaction of γ - FeOOH coated fibres (90% HFA-Cer, 10% S-Cer) with $\text{Fe}(\text{II})$ under N_2 to give magnetite (Fe_3O_4). c, Crystals of lepidocrocite grown between 6.5 nm lamellar lipid disks formed by acid conversion of unmineralized tubules (90% HFA-Cer, 10% S-Cer) before addition of $\text{Fe}(\text{III})$. Image is of an unstained sample. Intra-stack crystals (dark parallel lines) and sheet-like lepidocrocite overgrowths can be observed.

ration of sulphatide into the neutral galactocerebroside assemblies gives the microstructure anionic sites for electrostatic binding of Fe(III) polyhydroxy cationic species (which are formed at low pH²⁰). These binding sites are the initial foci for nuclei formation such that, under optimal structural and solution conditions, inorganic deposition is confined to the tubule surface. In the case of iron oxide mineralization, no preferential crystallographic orientation with respect to the lipid microstructure was observed, although this does not rule out the possibility of oriented nucleation with more crystalline minerals such as CaCO₃ and BaSO₄. Moreover, the presence of the iron mineral coating stabilizes the lipid microstructure with respect to hydration so that suspension of the high axial ratio composites can undergo chemical transformations; for example, to produce magnetic tubules. A range of inorganic-organic materials could conceivably be synthesized by this route. We have observed, however, that although amorphous calcium phosphate has a high nucleation affinity for S-Cer-doped HFA-Cer tubules, subsequent phase transformation to crystalline hydroxyapatite is often not associated with the lipid surface (D. Walsh and S.M., unpublished data). Thus, the interplay between interfacial and solution chemistry is critical in dictating the nucleation specificity of these inorganic-organic interfaces.

Finally, we note that although we have been successful in promoting inorganic nucleation on the external surface of the lipid microstructure, we have not yet been able to synthesize reproducibly tubule-based composites in which the inorganic phase resides solely within the 10–30 nm central lumen of the

supramolecular assembly. The synthesis of organic-lined inorganic filaments, as well as the production of viscoelastic gels of mineralized S-Cer-doped NFA-Cer unilamellar tubules (Fig. 1, route III) are the principal objectives of current and future work. □

Received 24 February 1993; accepted 14 June 1993.

1. Heuer, A. H. *et al. Science* **225**, 1098–1105 (1992).
2. Mann, S. *et al. MRS bulletin* **XVII**(10), 32–36 (1992).
3. Mann, S., Heywood, B. R., Rajam, S. & Birchall, J. D. *Nature* **334**, 692–695 (1988).
4. Zhao, X. K., Yang, J., McCormick, L. D. & Fender, J. H. *J. phys. Chem.* **96**, 9933–9938 (1992).
5. Gavish, M., Popovitz-Biro, R., Lahav, M. & Leiserovitz, L. *Science* **250**, 973–974 (1990).
6. Meldrum, F. C., Wade, V. J., Nimmo, D. L., Heywood, B. R. & Mann, S. *Nature* **349**, 684–687 (1991).
7. Meldrum, F. C., Heywood, B. R. & Mann, S. *Science* **257**, 522–523 (1992).
8. Mann, S., Hannington, J. P. & Williams, R. J. P. *Nature* **324**, 565–567 (1986).
9. Mendelson, N. H. *Science* **258**, 1633–1636 (1992).
10. Ekwali, P. *Adv. Liq. Cryst.* **1**, 1–142 (1975).
11. Yager, P. & Schoen, P. E. *Molec. Cryst. liq. Cryst.* **106**, 371–381 (1984).
12. Yamada, K., Ihara, H., Ide, T., Fukumoto, T. & Hirayama, C. *Chem. Lett.* 1713–1716 (1984).
13. Nakashima, N., Asakuma, S., Kim, J. M. & Kunitake, T. *Chem. Lett.* 1709–1712 (1984).
14. Fuhrop, J. H., Schneider, P., Rosenberg, J. & Boekema, E. J. *Am. chem. Soc.* **109**, 3387–3390 (1987).
15. Archibald, D. D. & Yager, P. *Biochemistry* **31**, 9045–9055 (1992).
16. Schnur, J. M. *et al. Thin Solid Films* **152**, 181–206 (1987).
17. Baral, S. & Schoen, P. *Chem. Mat.* **5**, 145–147 (1993).
18. Chappel, J. S. & Yager, P. *J. Mat. Sci. Lett.* **11**, 633–636 (1992).
19. Archibald, D. D. & Mann, S. *Chem. Phys. Lipids* (in the press).
20. Henry, M., Jolivet, J. P. & Livage, J. *Struct. and Bonding* **77**, 168–178 (1992).
21. Tamura, Y., Ito, K. & Katsura, T. *J. chem. Soc. Dalton Trans.* 189–194 (1983).

ACKNOWLEDGEMENTS. We gratefully acknowledge the financial support of the Science and Engineering Research Council.

Large igneous province on the US Atlantic margin and implications for magmatism during continental breakup

W. S. Holbrook & P. B. Kelemen

Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

RIFTED continental margins commonly include sections of igneous rock more than twice as thick as normal oceanic crust. Explanations for this voluminous magmatic accretion during rifting include plume models^{1–3}, which require a deep-seated thermal or chemical anomaly in upwelling mantle, and non-plume models^{4–7}, which call on broad, shallow thermal anomalies and/or rapid upwelling of mantle through the melting zone. New seismic models from two transects across the continent-ocean transition on the US Atlantic margin^{8–10} confirm the presence of a 20–25-km-thick igneous section. Here we argue that the similarity of the crustal structure on these and two previous transects, spanning 1,000 km of the margin, and the association of thick igneous crust with the East Coast magnetic anomaly¹¹ imply that the thick igneous section extends along the entire margin and may have a volume of as much as 3.2×10^6 km³. The distribution of volcanic and plutonic rocks, details of the seismic structure, and lack of independent evidence for a hotspot are difficult to reconcile with plume models and suggest that non-plume processes created the thick igneous crust.

Studies of the continental margins of the North Atlantic have shown that voluminous igneous activity, on a scale rivaling the largest continental flood basalt provinces, accompanied the onset of sea-floor spreading^{1,12}. Igneous activity on these 'volcanic margins' is shown in seismic data as seaward-dipping reflections beneath the post-rift unconformity on multichannel

seismic reflection (MCS) data^{13,14}, and thick, high-seismic-velocity (P-wave velocity 7.2–7.5 km s⁻¹) lower crust interpreted from wide-angle seismic data^{1,12}. Other margins (such as the Iberia margin), appear to be 'non-volcanic', as evidenced by lower seismic velocities (6.3–6.6 km s⁻¹) in the lower crust, lack of seaward-dipping reflections, and block-faulted basement interpreted as thinned continental crust^{15,16}.

The first indications of significant volumes of volcanic material in the crust of the eastern margin of the United States came from seaward-dipping reflections on MCS data¹⁷ and high velocity (7.2–7.5 km s⁻¹) lower crust interpreted from wide-angle seismic data in the Baltimore Canyon¹⁸ and Carolina Troughs¹⁹ (Figs 1, 2). Tréhu *et al.*¹⁹ inferred that the margin was volcanic. Due to limited resolution of those data sets, however, the full thickness and extent of igneous rocks was uncertain, and White and McKenzie²⁰ concluded that the margin was only mildly volcanic and formed at the distal edge of a hotspot.

Improved images of the geometry and distribution of rift-related igneous rocks have resulted from joint MCS and wide-angle seismic experiments conducted recently in the Carolina and southern Baltimore Canyon troughs^{8–10} (Fig. 1). The resulting images show well-developed seaward-dipping reflections together with a thick, high-velocity lower crust (Figs 3, 4). Beneath the seaward-dipping reflections, deep reflections from the Appalachian crust disappear, the Moho shows a striking change in reflective character (Fig. 3), and seismic velocities increase rapidly (Fig. 4). These abrupt lateral changes indicate that very little continental material persists beneath the volcanic wedge, so that the crust there consists almost entirely of new igneous material. The US Atlantic margin is therefore strongly volcanic at least from the Blake Spur fracture zone to offshore New Jersey, and probably along its entire length, as discussed below. The thickness of igneous material added to the crust reaches 25 km in the BA-6 and EDGE transects, exceeding the 18.5 km thickness interpreted on the Hatton Bank²¹, which was within 600–700 km of the Iceland hotspot during rifting. Although the inferred igneous material on the US Atlantic margin is buried beneath thick post-rift sediments and is therefore difficult to

sample directly, Early Jurassic mafic igneous rocks occur in the Newark-type basins and beneath the coastal plain²².

Comparison of magnetic anomaly data and the new seismic models shows a first-order spatial correlation between the thick volcanic crust and a broad zone of positive magnetic anomalies that includes the East Coast magnetic anomaly¹¹ (ECMA, Fig. 4). The landward limit of high-velocity igneous crust is marked by a sharp magnetic gradient near the basement hinge zone. Seaward, the magnetic intensity decreases across the transition from thick volcanic crust to oceanic crust of normal thickness. The peak of the ECMA lies between the thickest part of the seaward-dipping wedge and the landward limit of oceanic basement (Fig. 4). The offset between the thickest igneous section and the ECMA varies from profile to profile, suggesting that crustal magnetization varies laterally and extends into the lower crust^{8,23}. The relationship between the magnetic anomalies and inferred extent of thick igneous crust also holds on the LASE and USGS Line 32 profiles (Fig. 2).

We suggest that the similar crustal structure of the Carolina and Baltimore Canyon troughs, together with the association of

thick igneous crust with the ECMA, implies that the US Atlantic margin is strongly volcanic along the entire 2,000-km length of the ECMA, between 31° and 44° N (Fig. 1). We refer to the thick rift-related igneous package interpreted to underlie the margin as the East Coast Margin igneous province (ECMIP). The cross-sectional area of igneous material between rifted continental crust and normal oceanic crust is about 1,400 km² on transect BA-6 and 1,800 km² on EDGE 801. Taking 1,600 km² as an average area, we estimate the total volume of igneous material in the ECMIP as 3.2×10^6 km³. This is probably a maximum estimate, because it ignores the possible presence of intercalated sediments in the seaward-dipping wedge and relict continental crust within the lower crust, as well as possible gaps in the ECMIP. Even if the ECMIP does not extend much north of the LASE profile, the implied volume of 1.6×10^6 km³ classifies the ECMIP as a large igneous province, comparable in volume to the Deccan province (1.7×10^6 km³; ref. 24). The deep structure of the conjugate African margin is unknown, but if it is roughly symmetric with the North American margin, the total volume of material erupted during rifting may be twice our estimate.

Plume models have been invoked to explain several flood basalt provinces and volcanic margins. In its simplest form the hotspot model²⁰ predicts that the volume of igneous material emplaced in the crust depends on the temperature of the upper mantle during passive upwelling beneath rifting lithosphere. Rifting over mantle with a normal potential temperature (1,280 °C) produces igneous crust of normal oceanic thickness (7 km), whereas rifting over the warmer mantle near a hotspot produces greater thicknesses²⁰. Because the presence of water in the mantle also enhances the volume of melt produced during decompression, some 'hotspots' may be more 'wet' than 'hot'²⁵. Other plume models include a dynamic component. In 'plume-head' models the mantle upwelling rate varies with time: the flux of entrained, dynamically upwelling material is large during plume initiation, but decreases once the plume is established. Lateral migration and focusing of melt from a plume to the rift zone may increase igneous accretion without requiring temperature anomalies as large as 150–200 °C (refs 26, 27). Still, these models all require an externally imposed, deep-seated thermal or chemical anomaly in the asthenosphere to explain the formation of thick igneous crust.

Several lines of evidence suggest that the ECMIP was not produced by a long-lived plume. First is the lack of independent evidence for a plume near the margin during rifting. According to plate reconstructions²⁸ the closest likely hotspot to the central east coast during rifting was the Cape Verde hotspot, which was ~1,000 km northwest of the EDGE transect area 180 Myr ago, assuming a fixed hotspot reference frame. Even if we assume that the Cape Verde hotspot drifted in the necessary direction at 5 mm yr⁻¹, placing it under the margin near Long Island during rifting, there is no hotspot track of appropriate age and orientation. There is no chain of volcanoes, bathymetric swell, nor any known track of anomalously thick crust, leading eastward from the central east coast into the Atlantic. (The New England seamounts and Bermuda Rise are substantially younger than the opening of the margin, with ages of 124–75 Myr and 33 Myr, respectively²⁹.) A hotspot beneath North Africa might not have left a track on the Atlantic sea floor, but known African hotspots (for example, the Hoggar dome³⁰) were too far (at least 1,500 km) from the margin to have created the ECMIP. A 'hot blob' of rising material, lacking an underlying plume, is a possible explanation, but it requires a fortuitous arrival at the time of continental rifting.

In this respect the ECMIP may be similar to the Cuvier margin of northwestern Australia³¹ and is clearly different from margins where flood basalts and offshore igneous provinces have been linked to known plumes. In the South Atlantic, for example, the Tristan de Cunha hotspot, which is believed to have caused the Paraná basalts of South America and Etendeka basalts of Southwest Africa, left a distinct bathymetric track in the Walvis

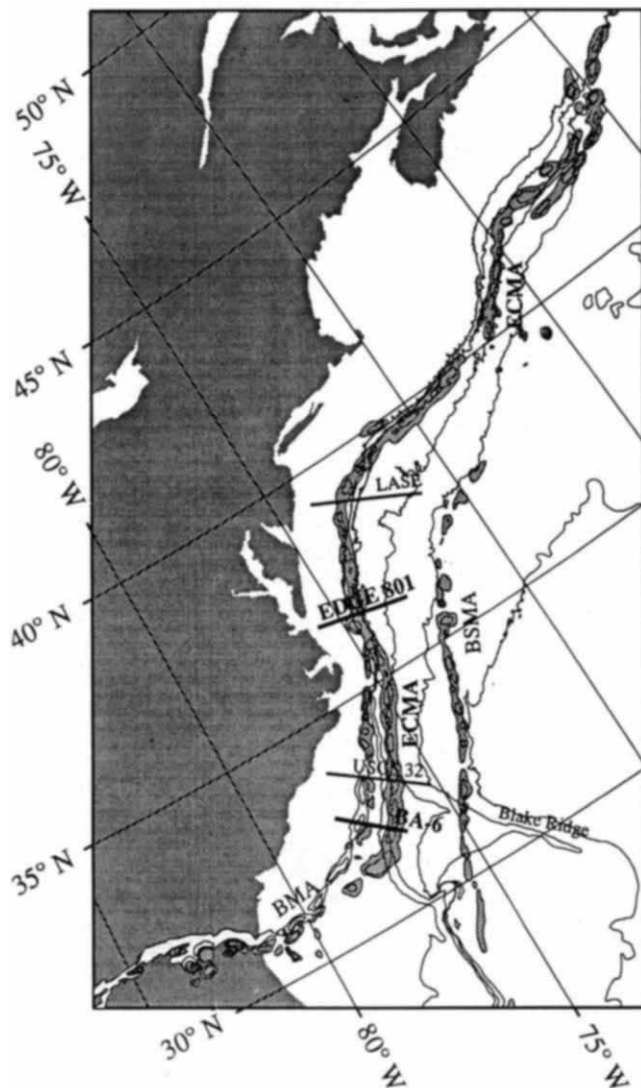
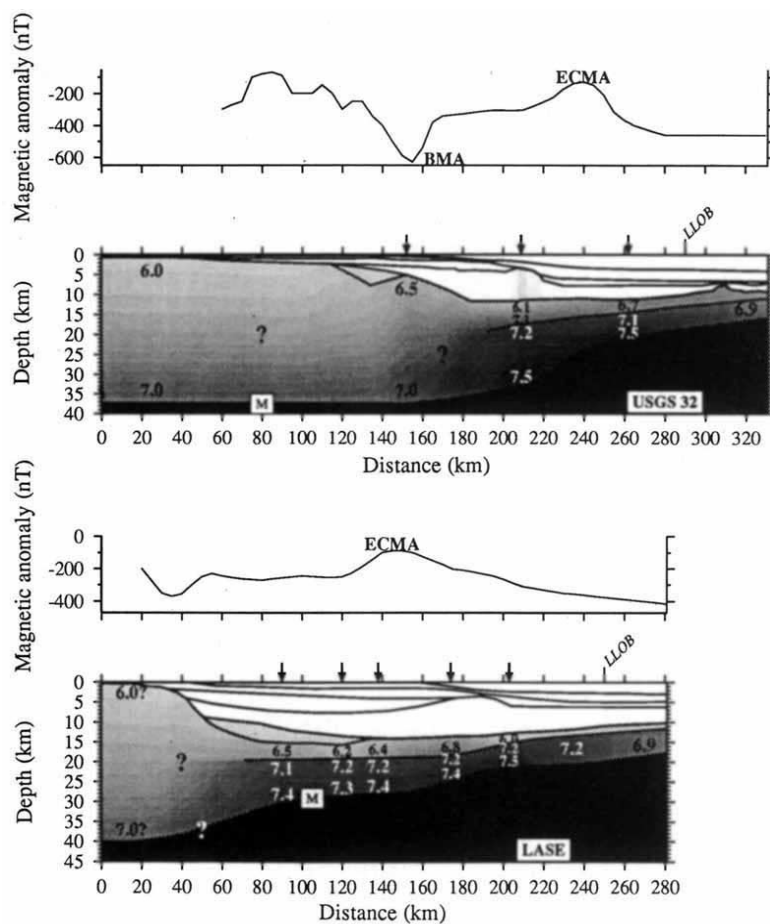


FIG. 1. Location of the US East Coast margin, bathymetry contoured at 1000 m intervals. Magnetic anomalies¹¹: ECMA = East Coast magnetic anomaly; BMA = Brunswick magnetic anomaly; BSMA = Blake Spur magnetic anomaly. Seismic transects that contain both reflection and refraction data: LASE¹⁸; EDGE Line 801⁹; Carolina Trough USGS line 32¹⁹; Carolina Trough line BA-6^{8,10}. The Blake Ridge is a sediment drift deposit³⁴.

FIG. 2 Magnetic data and crustal velocity structure across (top) USGS Line 32 transect, modified from ref. 19, and (bottom) LASE transect, modified from ref. 18. In the crustal velocity diagrams, shading is proportional to velocity and the numbers are velocities in km s^{-1} . Arrows indicate velocity control points from along-strike ocean-bottom seismometer lines or expanding spread profiles. M = Moho; ECMA = East Coast Magnetic Anomaly; BMA = Brunswick Magnetic Anomaly; LLOB = landward limit of oceanic basement, identified from coincident seismic reflection data.



ridge and Rio Grande rise. Similarly, the Iceland hotspot (North Atlantic Tertiary igneous province) left its mark on the sea floor in the Iceland–Faeroes and Greenland–Iceland rises, as did the Réunion hotspot (Deccan Traps) in the Chagos–Laccadive ridge. No such track exists in the central Atlantic.

Second, the strongly asymmetric distribution of igneous accretion along and across the margin is difficult to reconcile with plume models. The hotspot²⁰ and plume-head models^{2,3,32} imply that mantle temperature should decrease radially from the

plume. This in turn predicts that the thickness and magnesium content (and, therefore, seismic velocity) of igneous rocks will decrease radially from the plume. The similarity in thickness and velocity of igneous rocks on the BA-6 and EDGE transects, however, implies that melting did not vary strongly along the margin. Perpendicular to the margin, in contrast, the amount of melting decreased rapidly over a distance of only 60–80 km. From the thick igneous crust to oceanic crust, average igneous crustal velocities decrease from 7.1 to 6.4 km s^{-1} , the total thickness of igneous crust decreases from 25 to 7–8 km, and lower-crustal velocities decrease from 7.4–7.5 to 6.9–7.0 km s^{-1} (Fig. 4). This strong asymmetry in margin structure implies that igneous accretion is controlled by dynamic rift processes, rather than by proximity to a plume.

Third, the observed seismic velocities and crustal thicknesses are difficult to generate with the hotspot model without an unusually large thermal anomaly. The interpreted igneous crustal thickness of 25 km beneath the BA-6 and EDGE transects would require the maximum thermal anomaly (200 °C) considered reasonable by White and McKenzie. The observed lower-crustal velocities of 7.4–7.5 km s^{-1} cannot be generated with their model, which predicts a velocity of $\sim 7.15 \text{ km s}^{-1}$ for a 200 °C anomaly²⁰. Even if the lower crust comprises olivine-rich cumulates, the average igneous crustal velocities of 7.1 km s^{-1} require a 150 °C anomaly. It seems unlikely that such a high-temperature plume would not leave a hotspot track on the sea floor.

Thus, although plume models may explain some volcanic margins, we do not believe that they apply to the US Atlantic margin, and we must seek other explanations for the prodigious volcanism there. One possibility is that, during rifting, the upper mantle was warm over a broad region, without a deep-seated plume. Surface-wave tomography suggests that such 'hot cells' exist in the upper mantle, perhaps owing to a lack of subduction-related cooling⁶. Alternatively, upper-mantle thermal anomalies

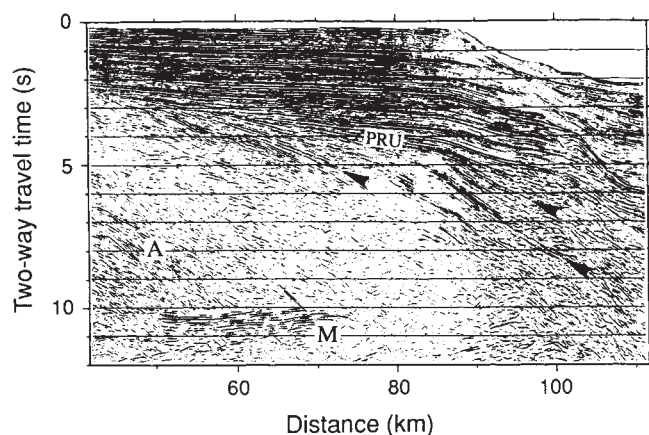


Fig. 3 Seismic reflection data showing seaward dipping reflections (arrows) on line EDGE 801. Strong, dipping lower-crustal reflections in Appalachian crust (A) give way seaward to a more transparent lower crust, whereas the Moho (M) shows high-amplitude laminations before disappearing beneath the thickest seaward-dipping reflections. These observations, together with the coincident velocity model (Fig. 4), indicate an abrupt transition from old continental crust to new igneous crust. PRU = post-rift unconformity.

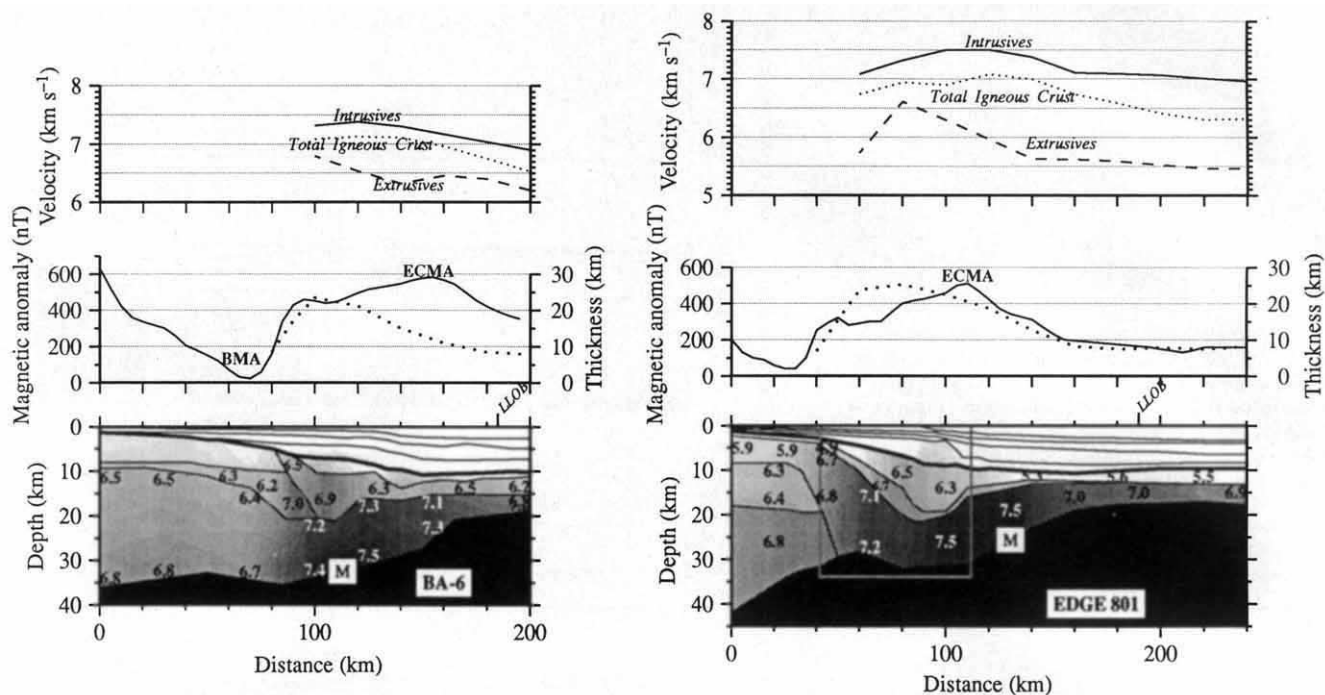


Fig. 4 Structure across Carolina Trough BA-6 profile³⁵ (left) and southern Baltimore Canyon Trough profile EDGE 801 (W.S.H. et al., manuscript in preparation) (right). Top, velocity across the section of (solid line) intrusives (lower crust), (broken line) extrusives (layer 2) and (dotted line) total igneous crust (extrusives + intrusives). Middle, Magnetic data in nT (solid line) and total thickness of igneous rocks emplaced during rifting, interpreted from seismic sections (dotted line). ECMA = East Coast magnetic anomaly; BMA = Brunswick magnetic anomaly.

maly. Bottom, Crustal velocity structure, with shading proportional to velocity; velocities labelled in km s^{-1} . Bold line represents post-rift unconformity. The strong lateral increase in crustal seismic velocity and first occurrence of seaward-dipping reflections correspond to a sharp seaward increase in magnetic intensity. Box outlines location of multi-channel seismic reflection data shown in Fig. 3. LLOB = landward limit of oceanic basement.

and consequent asthenospheric upwelling and continental rifting might result from the insulating effect of supercontinents⁵. The thermal anomalies envisioned in these models differ from deep-seated plumes in several ways: they are generated in the upper mantle, they originate over a broad region rather than at a point source, and they can be dissipated by subsequent rifting. Such thermal anomalies would not leave a distinct hotspot track.

We believe that dynamic mantle upwelling during rifting—in which the asthenospheric upwelling rate exceeds the lithospheric extension rate—contributes to the accumulation of thick igneous crust. Secondary convection, induced by temperature contrasts across an abrupt lithospheric rupture beneath the rift, may also enhance melting during continental breakup^{4,31,33}. Similarly, abrupt variations in the thickness of lithospheric domains juxtaposed by continental collision may lead to significant lateral temperature gradients which nucleate secondary convection before and during rifting⁷. The dynamics of these models, however, are poorly understood.

If dynamically upwelling mantle lacks a strong thermal anomaly, the liquid produced will initially consist of small melt fractions formed at high average pressure from a very large volume of peridotite. Such a process would explain the thick, high-velocity lower crust of the ECMIP. This hypothesis can be tested by geochemical analysis of new basalt samples from drilling of volcanic margins. We tentatively predict that such basalts will be found to be derived from relatively low-degree, high-pressure melting of the upper mantle, as found for basalts from the Vøring margin³³. □

6. Anderson, D. L., Zhang, Y. S. & Tanimoto, T. in *Magmatism at Continental Margins and the Causes for Continental Break-up*, Geol. Soc. Spec. Publ. no. 68 (eds Storey, B., Alabaster, T. & Pankhurst, R. J.) 99–124 (Geological Society of London, London, 1992).
7. Vogt, P. R. *Geology* **19**, 41–44 (1991).
8. Austin, J. A. Jr et al. *Geology* **18**, 1023–1027 (1990).
9. Sheridan, R. E. et al. *Geology* **21**, 563–567 (1993).
10. Holbrook, W. S., Reiter, E. C., Purdy, G. M. & Toksöz, M. N. *Geology* **20**, 203–206 (1992).
11. Klitgord, K. D. & Behrendt, J. C. *Mem. Am. Ass. Petrol. Geol.* **29**, 85–112 (1979).
12. Mutter, J. C., Talwani, M. & Stoffa, P. L. *J. geophys. Res.* **89**, 483–502 (1984).
13. Hinz, K. *Geol. Jb.* **22**, 5–28 (1981).
14. Mutter, J. C., Talwani, M. & Stoffa, P. L. *Geology* **10**, 353–357 (1982).
15. Whitmarsh, R. B., Miles, P. R. & Maufray, A. *Geophys. J. Int.* **103**, 509–531 (1990).
16. Pinheiro, L. M., Whitmarsh, R. B. & Miles, P. R. *Geophys. J. Int.* **109**, 106–124 (1992).
17. Klitgord, K. D., Hutchinson, D. R. & Schouten, H. in *The Geology of North America, Vol. 1–2, The Atlantic Continental Margin, U.S.* (eds Sheridan, R. E. & Grow, J. A.) 19–55 (Geological Society of America, Boulder, 1988).
18. LASE Study Group *Mar. Pet. Geol.* **3**, 234–242 (1986).
19. Tréhu, A. M. et al. *J. geophys. Res.* **94**, 10585–10600 (1989).
20. White, R. & McKenzie, D. *J. geophys. Res.* **94**, 7685–7729 (1989).
21. Morgan, J. V., Barton, P. J. & White, R. S. *Geophys. J. Int.* **98**, 367–384 (1989).
22. de Boer, J. Z., McHone, J. G., Puffer, J. H., Ragland, P. C. & Whittington, D. in *The Geology of North America, Vol. 1–2, The Atlantic Continental Margin, U.S.* (eds Sheridan, R. E. & Grow, J. A.) 217–241 (Geological Society of America, Boulder, 1988).
23. Talwani, M. et al. *EOS* **73**, 490–491 (1992).
24. Coffin, M. F. & Eldholm, O. in *Magmatism at Continental Margins and the Causes for Continental Break-up*, Geol. Soc. Spec. Publ. no. 68 (eds Storey, B., Alabaster, T. & Pankhurst, R. J.) 17–30 (Geological Society of London, London, 1992).
25. Bonatti, E. *Science* **250**, 107–111 (1990).
26. Vink, G. E. *J. geophys. Res.* **89**, 9949–9959 (1984).
27. Skogseid, J., Pedersen, T., Eldholm, O. & Larsen, B. T. in *Magmatism at Continental Margins and the Causes for Continental Break-up*, Geol. Soc. Spec. Publ. no. 68 (eds Storey, B., Alabaster, T. & Pankhurst, R. J.) 305–320 (Geological Society of London, London, 1992).
28. Morgan, W. J. in *The Sea, Vol. 7* (ed. Emiliani, C.) 443–487 (Wiley, New York, 1981).
29. Saemundsson, K. in *The Geology of North America, Volume M, The Western North Atlantic Region* (eds Vogt, P. R. & Tucholke, B. E.) 69–86 (Geological Society of America, Boulder, 1986).
30. Crough, S. T. *Tectonophysics* **77**, 189–202 (1981).
31. Hopper, J. R., Mutter, J. C., Larson, R. L., Mutter, C. Z. & Northwest Australia Study Group *Geology* **20**, 853–857 (1992).
32. Richards, M. A., Duncan, R. A. & Courtillot, V. E. *Science* **246**, 103–107 (1989).
33. Zehnder, C. M., Mutter, J. C. & Buhl, P. *Tectonophysics* **173**, 545–565 (1990).
34. Dillon, W. P. & Popenoe, P. in *The Geology of North America, Vol. 1–2, The Atlantic Continental Margin, U.S.* (eds Sheridan, R. E. & Grow, J. A.) 291–328 (Geological Society of America, Boulder, 1988).
35. Holbrook, W. S. et al. *J. geophys. Res.* (in the press).

Received 16 February 1993; accepted 1 June 1993.

1. White, R. S. et al. *Nature* **330**, 439–444 (1987).
2. Griffiths, R. W. & Campbell, I. H. *J. geophys. Res.* **96**, 18295–18310 (1991).
3. Campbell, I. H. & Griffiths, R. W. *Earth planet. Sci. Lett.* **99**, 79–93 (1990).
4. Mutter, J. C., Buck, W. R. & Zehnder, C. M. *J. geophys. Res.* **93**, 1031–1048 (1988).
5. Gurnis, M. *Nature* **332**, 695–699 (1988).

ACKNOWLEDGEMENTS. This research was supported by the NSF and by a WHOI Independent Study Award. We thank our collaborators, G. M. Purdy, M. Talwani, J. Austin, Jr and R. Sheridan, and acknowledge helpful discussions with C. Ruppel, K. Klitgord, U. ten Brink, and H. Schouten. We also thank J. Mutter for constructive comments on the manuscript.

Long-term tree populations in northwest Greece through multiple Quaternary climatic cycles

P. C. Tzedakis

Subdepartment of Quaternary Research,
Department of Plant Sciences, University of Cambridge,
Downing Street, Cambridge CB2 3EA, UK

THE emergent view of Quaternary cold stage European landscapes dominated by open vegetation communities has led to the concept that tree populations occurred only in restricted sites (refugia) in southern Europe¹⁻⁵, although there is still considerable uncertainty over the precise location and extent of such populations. Trees respond to Quaternary climatic change by spreading from refugia during interglacials, but at the end of an interglacial period there is no reverse migratory movement in the direction of refugia, as most northern populations degrade *in situ*. It has thus been proposed that survival of European trees through Quaternary climatic cycles is dependent on populations persisting continuously in southern Europe⁶. According to this model, it is these southern populations that furnish European interglacial forests with essentially the same components during the Quaternary and, moreover, it is failure to survive in these 'long-term refugia' that brings about a tree species' ultimate disappearance from Europe⁶. I present here a 430,000-year record of vegetational and climatic change from northwest Greece comparable to that of other long sequences, but exceptional in documenting the continuous presence of temperate tree pollen throughout the sequence. The levels and consistency of representation suggest the local occurrence of tree populations at fluctuating densities and distributions according to prevailing climatic regimes.

The investigation was carried out at Ioannina basin, northwest Greece (39° 40' N, 20° 51' E; 470 m above sea level), situated on the western flank of the Pindus Mountain Range (Fig. 1). The climate of the area is characterized by high levels of annual precipitation (1,200 mm) and a moderate summer drought⁷.

Ioannina is located within a region identified by palynological^{3,5,8,9} and physical⁶ evidence as likely to have supported refugial tree populations during cold stages. By comparison, the only Greek long pollen record hitherto available, Tenaghi Philippon¹⁰⁻¹³ to the east of the postulated refugial area (Fig. 1, inset), does not show any continuous presence of tree populations during glacial periods. A long core obtained by the Greek Institute of Geology and Mineral Exploration at the deepest part of the Ioannina basin was selected for palynological investigation.

Results show a higher order of magnitude changes in alternations between forest and open vegetation communities (Fig. 2a), a reflection of climatic shifts from interglacial to glacial modes. Superimposed on these oscillations is a lower order variability associated with vegetational changes within interglacial and glacial periods (Fig. 3)¹⁴. During forest periods a succession is recorded with *Quercus* and *Ulmus/Zelkova* populations expanding early, followed by *Carpinus betulus* and also *Ostrya carpinifolia*/*Carpinus orientalis*, and finally *Abies*. Individual forest periods often have a characteristic signature as particular taxa assume an overriding dominance: thus the lowermost forest complex has two separate phases characterized by high *Abies* values, the forest period above shows significant presence of *Abies* along with *Fagus* populations, the next major forest period (83.75–89.25 m) is dominated by *Quercus*, and the following period (45.88–59.00 m) has *Carpinus betulus* as its distinctive feature. The Holocene is not entirely represented here (ref. 8 has a complete record), and there are difficulties in distinguishing natural vegetation succession from anthropogenically related changes. Although relative abundances of trees may differ between periods, the underlying pattern of differential expansion remains distinct and consistent. Open vegetation periods also show a series of changes from transitional parkland, through grassland steppe communities, typically culminating in desert-steppe vegetation, representing an overall tendency towards more open, discontinuous vegetation during the course of a glacial period.

Correlation with Tenaghi Philippon^{10,11} and the marine oxygen isotope records provides a tentative chronology for the Ioannina 249 record (Fig. 3), indicating that the sequence down to a depth of 162.75 m represents a record of the past 423,000 years¹⁴. The sequence of events during the past 250,000 years as represen-

FIG. 1 Topographical map of the Ioannina area. Elevation of the Ioannina plain is 470 m; contour lines are at 200-m intervals starting at the 1,000 m level. The present size of Lake Pamvotis (22.8 km²) represents only part of its original extent, that stretched from Kato Lapsista to Katsikas, before it was drained this century. The bedrock geology is mainly comprised of Mesozoic limestones; Tertiary flysch is encountered in the southeastern edges of the basin. Location of borehole Ioannina 249 is shown as well as those of Ioannina I and II (ref. 8). Inset: simplified topographical map of southern Balkans: 0–400 m (unshaded), 400–1,000 m (dotted), and over 1,000 m (black).

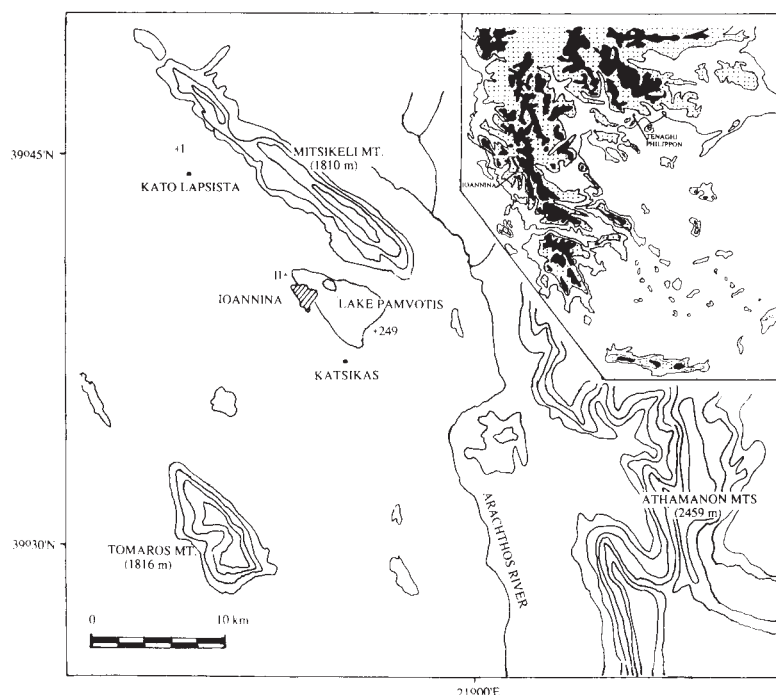


TABLE 1 Presence/absence of tree populations during forest and open vegetation periods at Ioannina

Depth (m)	Vegetation	<i>Pinus</i>	<i>Quercus</i>	<i>Ulmus</i>	<i>Corylus</i>	<i>Ostrya</i>	<i>Carpinus</i>	<i>Abies</i>	<i>Fagus</i>
10.20–17.25	F	?	+	+	+	+	+	+	(+)
17.25–33.25	O	(+)	+	(+)	?	–	–	+	(+)
33.25–36.75	F	+	+	+	+	(+)	+	+	+
36.75–40.25	O	+	+	–	–	–	–	+	–
40.25–44.13	F	+	+	+	+	+	+	+	+
44.13–45.88	O	+	+	+	+	+	+	+	+
45.88–59.00	F	+	+	+	+	+	+	+	–
59.00–83.75	O	+	+	+	(+)	–	?	+	–
83.75–89.25	F	–	+	+	+	+	+	+	+
89.25–112.75	O	(+)	+	–	–	–	–	?	–
112.75–117.00	F	+	+	+	+	+	+	+	(+)
117.00–139.00	O	+	+	+	(+)	(+)	+	+	–
139.00–143.25	F	+	+	+	+	+	+	+	+
143.25–151.00	O	+	+	+	(+)	(+)	+	+	?
151–162.75	F	+	+	+	+	+	+	+	(+)

Continuous percentage curves throughout a period are taken to reflect at least minimum abundance in the landscape, although adjustments need to take into account a taxon's tendency to be over- or under-represented and bias introduced by long-distance transport in spectra of low pollen-producing communities. *Pinus* and *Quercus* over-representation is a well known feature of pollen diagrams and is attributed to their high pollen productivity and dispersibility. Studies of modern pollen rain in samples from *Artemisia* steppe vegetation in the Iranian plateau²⁷, steppe regions in Syria²⁸, and western steppe, semi-desert and desert vegetation in the former USSR²⁹ show values for far travelled *Quercus* and *Pinus* pollen at or below 5% and 10%, respectively. Such values, however, should be treated with caution as they are derived mainly from moss polsters which are subject to the influence of local plants. As pollen of other temperate tree taxa is normally not recorded in such open vegetation environments, only generalized statements regarding their representation behaviour can be made here. *Abies* is heavily under-represented and has a restricted pollen dispersal area; modern pollen spectra from Greece record values up to 7.5% from *Abies*-dominated forest⁸. *Ulmus* is generally recorded in low pollen values in southern Europe⁵; in northwest Greece, when reported growing near a surface sampling site, its values in the pollen spectra reached only up to 3%³. In northwest Greece, *Corylus*, *Carpinus*, *Ostrya* and *Fagus* are today moderately-to-well represented in the pollen rain⁸. F, Forest period; O, open vegetation period; +, presence; – absence; (+), possible presence of populations at reduced abundances; ?, uncertainty between locally derived and far-travelled pollen. *Ulmus*, *Ulmus/Zelkova*; *Ostrya*, *Ostrya carpinifolia*/*Carpinus orientalis*; *Carpinus*, *Carpinus betulus*.

ted at Ioannina is also in general agreement with the record from Valle di Castiglione¹⁵, central Italy; changes in the primary structure of vegetation are similar, with *Fagus* and *Carpinus* having a more prominent role in Italian forests. The top 60 m of Ioannina 249, considered to contain the past 130,000 years, correlates well with the detailed records from Grande Pile^{16,17}, Les Echets¹⁸ and Plateau du Devès¹⁹. The last interglacial period followed by two interstadials is recorded at Ioannina in much the same fashion as the Eemian, St Germain I and II periods in France. During the full glacial, insufficient resolution at Ioannina 249 does not permit any detailed correlations with middle glacial interstadials of other records. In general, periods of open vegetation at the French sites are dominated by Gramineae, while *Artemisia* and chenopods are at higher abundances at Ioannina and Valle di Castiglione.

The nature of principal vegetation changes (forest to open vegetation) recorded in southern Europe and the relation between modern vegetation and climate in the Near East and further afield have led researchers to consider moisture availability as the critical climatic variable with temperature playing a supporting role^{15,20}. *Artemisia*- and chenopod-dominated pollen spectra have thus been considered to reflect increased aridity during the glacial periods. This however, is in conflict both with results from palaeoclimatic modelling of the Last Glacial Maximum (LGM) which show a steep January temperature decline and little change in precipitation²¹, or even increased seasonality of precipitation²², and also with geomorphological evidence for high LGM lake levels in the Mediterranean region²³ (including Ioannina²⁴). To address the conflict of higher lake levels and dry steppe vegetation, increased cloudiness and reduced evaporation rates in summer coupled with lowered winter temperature and precipitation levels have been invoked²³. Recently, water-balance and biome modelling applied to LGM conditions at Ioannina have suggested that increased winter precipitation may account for greater runoff and higher lake levels, while at the same time growing season soil moisture deficit and low winter temperatures would have maintained an open vegetation²⁵.

A prominent feature of the Ioannina record is the continuous presence of arboreal pollen during the open vegetation periods (Fig. 3): *Quercus* and *Pinus* pollen is normally recorded at values above 15 and 10%, respectively, and other trees are also represented but at lower abundances. Although the bias arising from long-distance pollen input of well dispersed taxa must be allowed for, the overall representation behaviour of trees in terms of consistency and replication of patterns is considered here the key aspect in deciphering local presence of tree populations during glacial periods. The continuous record of pollen of certain temperate trees (Fig. 3) provides evidence for the presence of small populations of these trees in the immediate vicinity of the Ioannina basin. The consistency of these patterns argues against representation biases arising from long-distance pollen transport and/or taphonomic processes. Evaluation of the Ioannina 249 record in terms of presence of temperate tree populations during forest and open vegetation periods (Table 1) shows the permanence of populations of *Quercus* and *Abies* in the landscape, followed by *Pinus*, *Ulmus/Zelkova* and, to a certain extent, *Corylus*. These populations reach highest densities during interglacial periods but continue to persist at low abundances during glacial periods. The less consistent representation of other tree taxa is difficult to assess, but their early interglacial appearance may indicate relative proximity of refugial locations, or alternatively survival in the immediate vicinity at population densities below palynological detection⁶.

By comparison, pollen of temperate trees during open vegetation periods in other long sequences (Tenaghi Philippon^{10,13}; Valle di Castiglione¹⁵ and French sites^{16–19}) is either recorded intermittently (for over-represented taxa) or mostly absent (for under-represented taxa). High precipitation levels and topographical variability (Fig. 1) at Ioannina are factors that may account for the local presence of refugial populations compared to other sites. Modelling results²⁵ showing increased winter precipitation at Ioannina may support such explanation, if some moisture accumulated in the soil could be available for the spring growing season. Note, however, that although such experi-

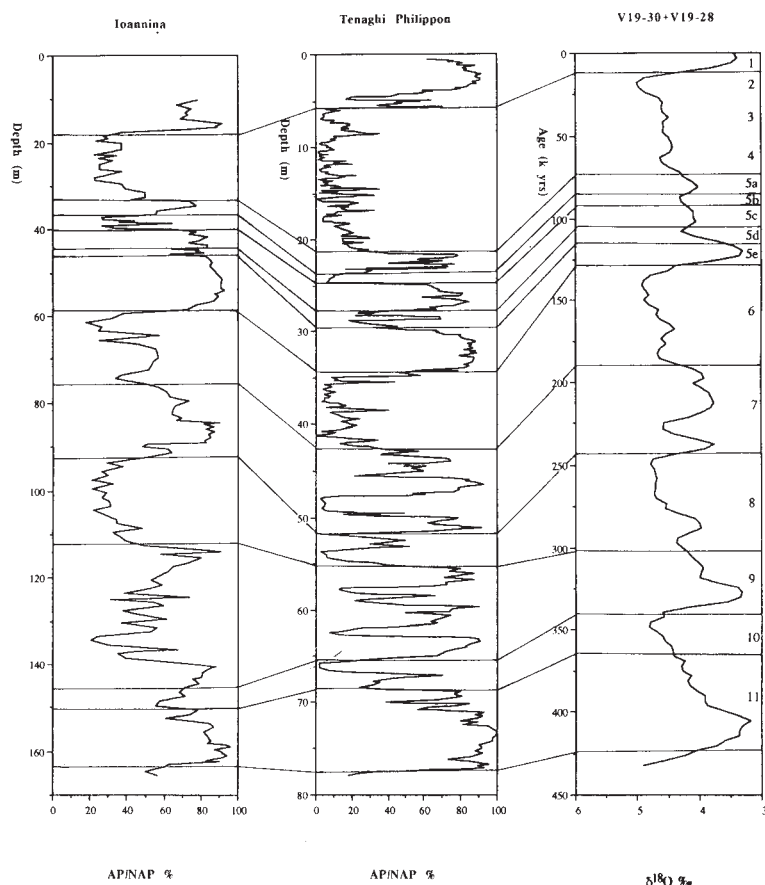
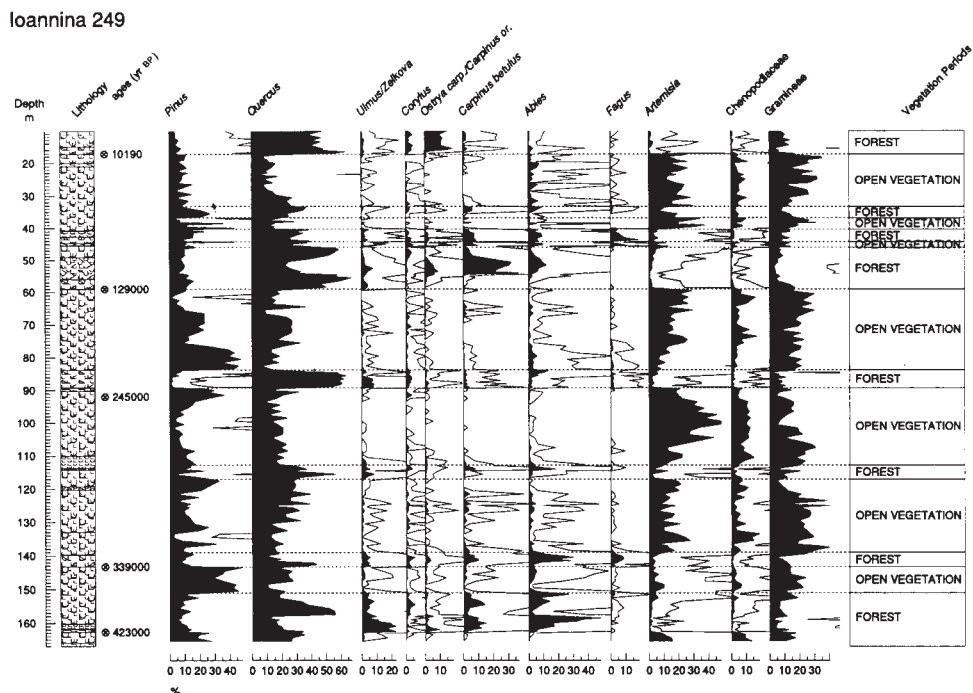


FIG. 2 Summary arboreal pollen (AP)/non-arboreal pollen (NAP) percentages plotted against depth for Ioannina 249 and Tenaghi Philippon (the latter are redrawn from refs 10, 11); note different depth scales. Also shown is the composite oxygen isotope record of V19-30 (3° 21' S, 83° 21' W)³⁰ for the interval 0–340 kyr BP, and V19-28 (2° 22' S, 84° 39' W)³¹ for 340–432 kyr BP, plotted against time isotopic stages and substages are indicated. Proposed correlations between the different records are presented. Correlations between Ioannina 249 and Tenaghi Philippon were based on the primary structure of vegetational development and relative stratigraphical position of different periods¹⁴. Correlation of Ioannina 249 with the marine oxygen isotope record was done using Tenaghi Philippon as an intermediate 'stepping stone', because Tenaghi Philippon has an age model developed from a series of radiocarbon dates from the upper part of the sequence, a generally accepted approximation of 120 kyr BP for the last interglacial and the age of the Brunhes/Matuyama magnetic reversal boundary³². On the basis of its independent chronology, Tenaghi Philippon has been correlated with the oxygen isotope record³².

FIG. 3 Pollen percentage diagram of selected taxa. Calculation sum includes pollen of terrestrial vascular plants. Ten times exaggeration is also shown. Total pollen counts ranged from 200 to 1,033 (mean 413). Nine samples had counts between 200 and 300 because of low pollen concentration in glacial age sediments. Sampling interval ranged from 25 cm to 100 cm. Lithostratigraphic column is presented; symbols are based on ref. 33. Sediments are grey silts having fine sand component in places and with varying calcium carbonate content¹⁴. Sediments of interglacial age have increased calcium carbonate and organic content, whereas sediments of glacial age are generally inorganic. Preservation of pollen grains in interglacial age sediments was good; grains in the inorganic glacial-age sediments showed a greater degree of deterioration which in most cases did not inhibit identification. Also shown are ages assigned to major transitions from open vegetation to forest periods based on: (1) correlation with Ioannina I (ref. 8) for the late glacial Holocene transition; (2) correlation with the oxygen isotope record; the chronology used was that of refs 34, 35. Only ages for the glacial-to-interglacial transitions (or terminations) corresponding to open vegetation-to-forest transitions are used here. The operative



assumption is the synchronicity of the events in the oxygen isotope record and in the terrestrial vegetation record which may be considered reasonable when applied to terminations which are relatively rapid and large scale events. Note that the uppermost age is in radiocarbon years, whereas the others are in calendar years.

ments have considered Ioannina as a typical case for the northern Mediterranean region²⁵, the pollen evidence presented here suggests that it is in fact an exceptional site and that any palaeoclimatic conclusions should be considered particular to this area.

Although individual forest periods during the Quaternary in Europe may be characterized by dominance of particular taxa, their species composition and vegetation succession are remarkably similar. This may be a reflection of the permanence of southern populations that supply northern forests with essentially the same components every climatic cycle. Despite the persistence of the southern populations, speciation has not been a feature of the Quaternary: any microevolutionary change that may have accumulated during a glacial period in isolated popu-

lations is lost as communities are reorganized during interglacials²⁶. Changes in species composition occur when inability to survive in refugia leads to extinction⁶. The evidence of permanence of at least certain tree populations at Ioannina through multiple interglacial as well as glacial periods lends support to the notion of 'long-term refugia' and underlines the importance of present southern European tree populations for the maintenance of tree species in Europe on the Quaternary timescale. The pressures exerted on southern populations by increasing global aridity and anthropogenic practices during the course of the present interglacial may thus lead many European tree species to extinction by the end of the next climatic cycle and result in substantially modified future interglacial vegetation. □

Received 5 March; accepted 19 May 1993.

1. Frenzel, B. *Science* **161**, 637–649 (1968).
2. Lang, G. *Feddes Rep.* **81**, 315–335 (1970).
3. Van der Hammen, T., Wijmstra, T. A. & Zagwijn, W. H. in *The Late Cenozoic Glacial Ages* (ed. Turekian, K. K.) 391–424 (Yale Univ. Press, New Haven, 1971).
4. Beug, H.-J. *Biul. Geol. Warszawa* **19**, 101–110 (1975).
5. Huntley, B. & Birks, H. J. B. *An Atlas of Past and Present Pollen Maps for Europe; 0–13,000 Years Ago* (Cambridge Univ. Press, Cambridge, 1983).
6. Bennett, K. D., Tzedakis, P. C. & Willis, K. J. *J. Biogeogr.* **18**, 103–115 (1991).
7. National Statistical Service of Greece *Statistical Yearbook of Greece* (Athens, 1951–1980).
8. Bottema, S. thesis, Univ. Groningen (1974).
9. Willis, K. J. *New Phytol.* **121**, 101–117, 119–138, 139–155 (1992).
10. Wijmstra, T. A. *Acta bot. Neerl.* **18**, 511–527 (1969).
11. Wijmstra, T. A. & Smit, A. *Acta bot. Neerl.* **25**, 297–312 (1976).
12. Van der Wiel, A. M. & Wijmstra, T. A. *Rev. Palaeobot. Palynol.* **52**, 73–88 (1987).
13. Van der Wiel, A. M. & Wijmstra, T. A. *Rev. Palaeobot. Palynol.* **52**, 89–117 (1987).
14. Tzedakis, P. C. thesis, Univ. Cambridge (1991).
15. Follieri, M., Magri, D. & Sadori, L. *Pollen Spores* **30**, 329–356 (1988).
16. Woillard, G. *Quat. Res.* **9**, 1–21 (1978).
17. Woillard, G. & Mook, W. *Science* **215**, 159–161 (1982).
18. Beaulieu, J.-L. de & Reille, M. *Boreas* **133**, 111–132 (1984).
19. Pons, A., Guiot, J., Beaulieu, J.-L. de & Reille, M. *Quat. Sci. Rev.* **11**, 439–448 (1992).
20. Bottema, S. in *Evaluation of Climate Proxy Data in Relation to the European Holocene* (ed. Frenzel, B.) 64–79 (Akademie der Wissenschaften und der Literatur, Mainz; European Science Foundation, Strasbourg; Gustav Fischer, Stuttgart, 1992).

21. Kutzbach, J. E. & Guetter, P. J. *J. Atmos. Sci.* **43**, 1726–1759 (1986).
22. Broccoli, A. J. & Manabe, S. *Clim. Dynam.* **1**, 87–99 (1987).
23. COHMAP Members *Science* **241**, 1043–1052 (1988).
24. Higgs, E. S., Vita-Finzi, C., Harris, D. R. & Fagg, A. E. *Proc. Prehist. Soc.* **33**, 1–26 (1967).
25. Prentice, I. C., Guiot, J. & Harrison, S. P. *Nature* **360**, 658–660 (1992).
26. Bennett, K. D. *Paleobiology* **16**, 11–21 (1990).
27. Wright, H. E., Jr, MacAndrews, J. H. & van Zeist, W. J. *Ecol.* **5**, 415–443 (1967).
28. Zeist, W. van & Bottema, S. *Late Quaternary Vegetation of the Near East* (Reichert, Wiesbaden, 1991).
29. Peterson, G. M. *Quat. Sci. Rev.* **2**, 281–321 (1983).
30. Shackleton, N. J. & Pisias, N. G. in *The Carbon Cycle and Atmospheric CO₂: Natural Variations Archean to Present* (eds Sundquist, E. T. & Broecker, W. S.) 303–317 (American Geophysical Union, Washington DC, 1985).
31. Ninkovich, D. & Shackleton, N. J. *Earth planet. Sci. Let.* **27**, 20–34 (1975).
32. Wijmstra, T. A. & Groenhard, M. C. *Rev. Acad. Colomb. Exactas, Fisi. Nat.* **15**, 87–98 (1984).
33. Troels-Smith, J. *Dan. Geol. Unders. IV Raekke* **3**, 1–73 (1975).
34. Martinson, D. G. et al. *Quat. Res.* **27**, 1–29 (1987).
35. Imbrie, J. et al. in *Milankovitch and Climate* (eds Berger, A., Imbrie, J., Hays, J., Kukla, G. & Saltzman, B.) 269–306 (Reidel, Dordrecht, 1984).

ACKNOWLEDGEMENTS. I thank K. D. Bennett, D. Magri and K. J. Willis for comments on the manuscript and helpful discussions, the staff of the Institute of Geology and Mineral Exploration in Athens and Ioannina for making available the cores from Ioannina and for their invaluable co-operation, and N. J. Shackleton for providing oxygen isotope data. The work was supported by a NERC research studentship.

Confidence in evolutionary trees from biological sequence data

M. A. Steel*, P. J. Lockhart† & D. Penny†

* Department of Mathematics, † Molecular Genetics Unit, Massey University, Palmerston North, New Zealand

THE reliable construction of evolutionary trees from nucleotide sequences often depends on randomization tests such as the bootstrap¹ and PTP (cladistic permutation tail probability) tests^{2,6}. The genomes of bacteria⁷, viruses⁸, animals^{7,9,10} and plants¹¹, however, vary widely in their nucleotide frequencies. Where genomes have independently acquired similar G+C base compositions, signals in the data arise that cause methods of evolutionary tree reconstruction to estimate the wrong tree by grouping together sequences with similar G+C content^{12–14}. Under these conditions randomization tests can lead to both the rejection of the correct evolutionary hypothesis and acceptance of an incorrect hypothesis (such as with the contradictory inferences from the photosynthetic *rbcS* and *rbcL* sequences¹⁴). We have proposed one approach to testing for the G+C content problem¹⁵. Here we present a formalization of this method, a frequency-dependent significance test, which has general application.

Suppose we have a collection of four sequences of r -state characters (in practice, r will usually be 2, 4 or 20). If we edit the sites so that only the parsimony sites (phylogenetically informative sites) are present, then c will denote the length of these (edited) sequences. Let T_1 , T_2 and T_3 be the three binary trees on the four taxa, where T_i is the tree in which taxon 1 and taxon $i+1$ are grouped together. Let n_i denote the number of sites that

favour tree T_i , that is, which fit on T_i with a single mutation. Consider the following null model, in which four sequences are generated purely randomly (each state is independently chosen) with the proviso that the expected frequencies of the states for each sequence equal the actual observed frequencies. Then, by chance alone, among c informative sites generated in this way, a certain random number, which we denote as N_i , will favour tree T_i . We now calculate the probability $Pr[N_i \geq n_i]$ that this random contribution is at least the observed value n_i . We also describe a way of normalizing n_i to take account of this random contribution. Let $\rho_i(j)$ denote the proportion of sites in sequence $i=1 \dots 4$ at which state $j=1 \dots r$ appears. Evaluate the three sums

$$S_1 = \sum_{j \neq k} \rho_1(j) \rho_2(j) \rho_3(k) \rho_4(k);$$

$$S_2 = \sum_{j \neq k} \rho_1(j) \rho_2(k) \rho_3(j) \rho_4(k);$$

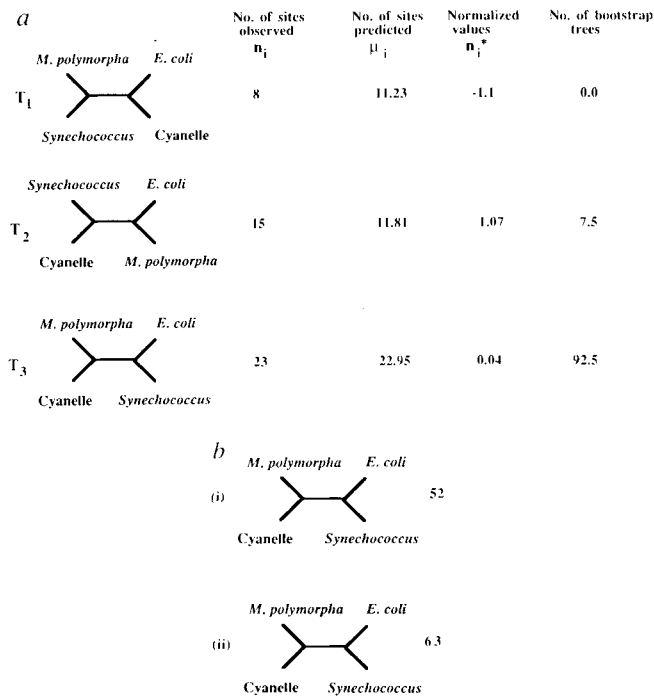
$$S_3 = \sum_{j \neq k} \rho_1(j) \rho_2(k) \rho_3(k) \rho_4(j),$$

and let

$$s_i = \frac{S_i}{S_1 + S_2 + S_3} \quad \text{for } i = 1, 2, 3.$$

Then under the null model, the number, N_i , of sites that fit tree T_i in a random sample of c informative sites has a binomial distribution $B(c, s_i)$. Thus, under the null model, the probability that N_i is at least as large as its observed value, n_i , is given by the cumulative binomial value

$$Pr[N_i \geq n_i] = \sum_{k \geq n_i} \binom{c}{k} s_i^k (1-s_i)^{c-k} \quad (1)$$



Note that N_i has a mean and standard deviation given by $\mu_i = cs_i$, $\sigma_i = \sqrt{cs_i(1-s_i)}$. Thus, under the null model the normalized value $n_i^* = (n_i - \mu_i)/\sigma_i$ has mean 0 and variance 1. Also, if c is large, n_i^* has, approximately, a standard normal distribution under the null model, which allows the rapid estimation of equation (1) when $c > 50$. Whatever the size of c , it seems sensible to compare trees not on n_i , the number of sites favouring tree T_i , but on the normalized values n_i^* . Also, it is possible to test the null model itself by computing the statistic

$$\chi^2 = \sum_{i=1}^3 \frac{(n_i - \mu_i)^2}{\mu_i} \quad (2)$$

Under the null model, and provided each μ_i is not too small (say ≥ 5) this statistic has a χ^2 distribution with two degrees of freedom. Equations (1) and (2) provide a frequency-dependent significance test (FD test) for the data under the null model.

Note that, for nucleotide sequences, our null model allows a limited (random) degree of variation in nucleotide frequencies between distinct regions or subsets of sites (for example, coding/non-coding regions or triplet positions). But if these regional

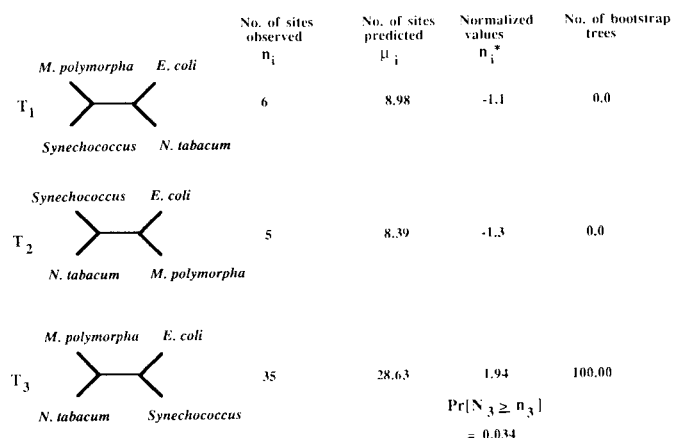
FIG. 1 **a**, The three possible unrooted binary trees for a four-taxon case. Analysis was on the 46 parsimony sites at the first and second codon positions. The cyanelle and chloroplast sequences are used as part of a study on the endosymbiotic origin of organelles and here relationships are difficult to assess^{14,15,18} as organelles generally have a high A + T content. The frequencies of the bases are cyanelle (0.239a, 0.174c, 0.174g, 0.413t), *E. coli* (0.261a, 0.478c, 0.196g, 0.065t), *Synechococcus* (0.217a, 0.391c, 0.283g, 0.109t) and *Marchantia polymorpha* (0.196a, 0.261c, 0.174g, 0.370t). The number of sites supporting each tree (n_i) is compared with the expected number (μ_i) under the null model. The normalized values n_i^* are given. None of the three values $Pr[N_i \geq n_i]$ is significant even at the 0.1 level (the FD test described by equation (2) gives $\chi^2 = 1.79$, d.f. = 2). Applying the PTP test, 8 out of 100 (column) randomized data sets produced trees that under parsimony were shorter or equal to that constructed from the original data, indicating significance for T_3 . Bootstrap analysis also gave spurious support to T_3 (92.5/100 recovered). **b**, Majority-rule consensus trees for (row, not column) randomized sequences (i, upper) singleton and parsimony sites at first and second codon positions: 284 sites, and (ii, lower) parsimony sites: 46 sites. Sequences were randomized 100 times and trees reconstructed from these randomized sequences; the values shown are the number of times the given tree was supported. The expected number for the three possible unrooted trees, given a symmetric base composition, would be 33/100. The results are significantly different from 33, showing that the tree-building method grouped the sequences of similar G + C content, even when the sequences were randomized.

nucleotide frequencies differ significantly, then a modified null model, in which sequences are generated randomly but subject to the regional frequency constraints, would predict different numbers of sites supporting each tree from those predicted by our null model (which is based on averages of nucleotide frequencies). For sites from a specified sequence, where varying G + C content within a genome is not important for the sequence, and with constant sites edited out, this factor is not important (this applies for our data sets). In cases where this factor is important, the calculations described below should be carried out separately on the regions and combined using standard statistical techniques.

We illustrate our test for two data sets; one in which the 'historical signal' (patterns in the data that indicate the true species relationship and which have not arisen from convergence) is expected^{14,15} to have been lost through multiple substitutions, and a second data set in which we expect some phylogenetic information still to be present.

Figure 1a shows three binary unrooted trees for *atpA* sequences from four taxa. This gene encodes the α -subunit of ATP synthetase and is highly conserved in the genomes of mito-

FIG. 2 The three unrooted binary trees for a four-taxon case in which a historical signal is expected between the chlorophyll *a/b* containing taxa (replacing the cyanelle sequence with a sequence from tobacco (*Nicotiana tabacum*)). Analysis was carried out on parsimony sites at the first and second codon positions considered jointly (also 46 sites). For tree T_3 , we have $n_3 = 35$, $\mu_3 = 28.63$, and from equation (1), $Pr[N_3 \geq n_3] = 0.034$, so according to the FD test, T_3 is significant at the 0.05 level. The frequencies of the bases were *Nicotiana tabacum* (0.326a, 0.217c, 0.087g, 0.370t), *E. coli* (0.174a, 0.543c, 0.261g, 0.022t), *Synechococcus* (0.174a, 0.478c, 0.217g, 0.130t) and *Marchantia polymorpha* (0.370a, 0.152c, 0.087g, 0.391t). Applying the PTP test, 0 (out of 100) randomized data sets produced trees that under parsimony were shorter or equal to that constructed from the original data, thus supporting the original tree as significant. The bootstrap test also gave strong support to the expected tree: T_3 (100/100 recovered). Note that the FD test could be extended to more than four taxa, either for parsimony or compatibility, although calculation of the s_i values becomes more difficult as the number of taxa grows (M.A.S. and P.J.L., manuscript in preparation).



chondria, chloroplasts, photosynthetic and non-photosynthetic eubacteria. The sequences are from the cyanelle (a photosynthetic organelle, characterized by a pycobilin light-harvesting system and found in the protist *Cyanophora paradoxa*), *Marchantia polymorpha* chloroplast (characterized by a chlorophyll *a/b* light-harvesting complex), *Synechococcus* PCC 6301 (a cyanobacterium, characterized by a phycobilin light-harvesting system) and *Escherichia coli* (a non-photosynthetic eubacterium). The edited sequences have a G+C content of 35, 43, 67 and 67%, respectively. The number of parsimony sites n_i that support the three possible trees is given. Greatest support is for tree T_3 , and bootstrap support for T_3 is also strong (92.5/100 trees). Similarly, analysis of the data using a PTP (cladistic permutation tail probability) test⁵ indicates significance at the 0.1 level, in that only 8 trees out of the 100 (column-randomized) data sets were shorter than or equal to the original tree.

In contrast, application of our FD test shows that the patterns observed in the data are close to random. The reason for this disparity is that similar G+C contents may suggest phylogenetic structure even when there is none. To emphasize this, we can randomize each sequence across the variable or parsimony sites and construct trees from these (row) randomized sequences. Usually, the original tree will be recovered (Fig. 1b), even though the sequences are random, apart from their G+C content.

This example illustrates an important but poorly appreciated distinction in taxonomy between convergence (the degree to which longer sequences (less sampling error) would result in the same tree) and consistency (the increasing certainty that the tree generated from the sequences is the correct tree—the tree that generated the sequences). In the present example both the PTP and bootstrap results suggest a high level of convergence, but because the sequences are effectively random, no tree building method can be consistent.

Figure 2 gives a second example in which the cyanelle sequence has been replaced with a chloroplast sequence from tobacco (*Nicotiana tabacum*). In this case, greatest parsimony support is also for T_3 . The tree is strongly supported by the bootstrap (100/100 trees) and the PTP test. With this data set our FD test indicates non-random patterns (presumably historical patterns) shared between tobacco and liverwort (*Marchantia polymorpha*), thus giving confidence that the tree is not an artefact of G+C composition.

For some data sets, there could be many determinants of pattern that result in contradictory signals. Irregular base composition is one such determinant which might converge sequences^{12–14,17} both at silent and replacement sites in coding genes, as well as in ribosomal RNAs^{7,8,10,13,14,16}. In view of this, our test should help to evaluate whether near or closest neighbours in phylogenetic trees share 'historically informative patterns', or whether they group together solely because of similar base composition. □

Received 24 February; accepted 14 May 1993.

1. Felsenstein, J. *Evolution* **39**, 783–791 (1985).
2. Archie, J. W. *Syst. Zool.* **38**, 239–252 (1989).
3. Henderson, I. M., Penny, D. & Hendy, M. D. *Nature* **326**, 22 (1987).
4. Henderson, I. M., Hendy, M. D. & Penny, D. *J. theor. Biol.* **140**, 289–303 (1989).
5. Faith, D. & Cranston, P. S. *Cladistics* **7**, 1–28 (1991).
6. Steel, M. A., Hendy, M. D. & Penny, D. *J. Class.* **9**, 71–90 (1992).
7. Jukes, T. H. & Bhushan, V. *J. molec. Evol.* **24**, 39–44 (1986).
8. Keese, P., MacKenzie, A. & Gibbs, A. *Virology* **172**, 536–546 (1989).
9. Bernardi, G. & Bernardi, G. *J. molec. Evol.* **24**, 1–11 (1986).
10. Crozier, R. H. & Crozier, Y. C. *Genetics* **133**, 97–117 (1993).
11. Oliver, J. L., Marin, A. & Martinez-Zapater, J. M. *Nucleic Acids Res.* **18**, 65–73 (1990).
12. Penny, D., Hendy, M. A., Zimmer, E. A. & Hanby, R. K. *Aust. Syst. Bot.* **3**, 21–38 (1990).
13. Lockhart, P. J., Howe, C. J., Bryant, D. A., Beanland, T. J. & Larkum, A. W. D. *J. molec. Evol.* **34**, 153–162 (1992).
14. Lockhart, P. J. et al. *FEBS Lett.* **301**, 127–131 (1992).
15. Lockhart, P. J. & Penny, D. *Research in Photosynthesis* Vol. III, 499–505 (Kluwer, Dordrecht, 1992).
16. Hasegawa, M. & Hashimoto, T. *Nature* **361**, 23 (1993).
17. Saccone, C., Pesole, G. & Preparata, G. *J. molec. Evol.* **29**, 407–411 (1989).
18. Lockhart, P. J., Penny, D., Hendy, M. D. & Larkum, A. W. D. *Photosyn. Res.* (in the press).

ACKNOWLEDGEMENTS. Supported by a grant from the New Zealand Lotteries Commission (M.A.S.) and the New Zealand Health Research Council (P.J.L.).

Antagonistic chromatic mechanisms in photoreceptors of the parietal eye of lizards

Eduardo Solessio & Gustav A. Engbreton

Institute for Sensory Research, and Department of Bioengineering and Neuroscience, Syracuse University, Syracuse, New York 13244, USA

PHOTORECEPTORS are the first in the chain of neurons that process visual information. In lateral eyes of vertebrates, light hyperpolarizes rod and cone photoreceptors that synapse onto bipolar and horizontal cells in the first synaptic layer of the retina. The sign of the photoreceptor signal is either conserved or inverted in bipolar cells, resulting in chromatically dependent depolarizing and hyperpolarizing responses to visual stimuli. Visual information is then conveyed to the second synaptic layer for encoding and transmission to the brain by ganglion cells. The parietal (third) eye of lizards does not contain bipolar cells or other interneurons. Photoreceptors synapse directly onto ganglion cells^{1–4} and yet, even in the absence of interneurons, antagonistic chromatic mechanisms modulate the ganglion cell responses^{5,6}. We report here that chromatic antagonism in the third eye originates in the chromatically dependent hyperpolarizing and depolarizing response of the photoreceptors to light. We also suggest that the antagonistic nature of these photoresponses may provide lizards with a mechanism for the enhanced detection of dawn and dusk.

The parietal eye of lizards is a simple yet highly structured photoreceptive organ (Fig. 1) that projects to non-visual areas in the midbrain^{7,8}. Intracellular recordings from photoreceptors in the parietal eye reveal the existence of antagonistic chromatic interactions already at this level: that is, the photoreceptors depolarize in response to a green light and hyperpolarize in response to a superimposed blue stimulus (Fig. 2).

We recorded intracellularly from photoreceptors of excised parietal eye preparations from *Xantusia vigilis* and *Uta stansburiana*. Typical membrane potentials in the dark were -50 mV and green flashes elicited depolarizations of up to 20 mV. Intense stimuli produce a characteristic transient at onset (Fig. 2a). The responses are graded with stimulus intensity and the maximal spectral sensitivity of the dark-adapted photoreceptors is 495 nm (Fig. 2e). The responses to shorter-wavelength stimuli differ somewhat from the green-driven response, especially at higher stimulus intensities. Responses to intense blue flashes lack the pronounced initial transient of green-driven responses but produce an additional depolarization on cessation of the stimulus (Fig. 2b). Intensity–response curves reveal that the responses to blue and green stimuli grow similarly at lower stimulus intensities (Fig. 2d), which is evidence that they are produced by a common chromatic mechanism. But at higher intensities the intensity–response functions diverge. The green intensity response function saturates like that of other vertebrate photoreceptors⁹, whereas that of the blue response reaches a maximum at lower intensity levels and then decreases with increasing stimulus intensity. This suggests that a second antagonistic chromatic mechanism is activated by intense blue stimuli.

Chromatic adaptation with 546 -nm light uncovered a blue-sensitive mechanism with maximum sensitivity at 440 nm (Fig. 2e). Presentation of the green adapting background depolarizes the membrane and superimposed blue stimuli then cause graded hyperpolarizations (Fig. 2c), driving the membrane potential back towards its original resting potential. The hyperpolarizing response is less sensitive (~ 2.0 log units) than the long-wavelength depolarizing response. Injection of Lucifer yellow or neu-

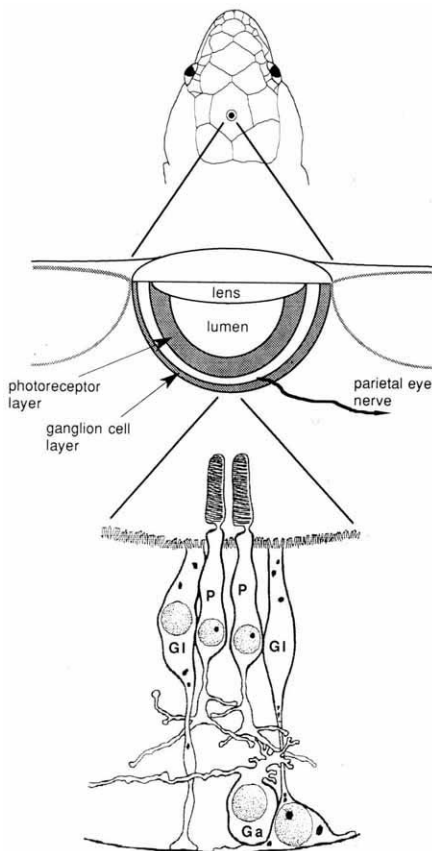
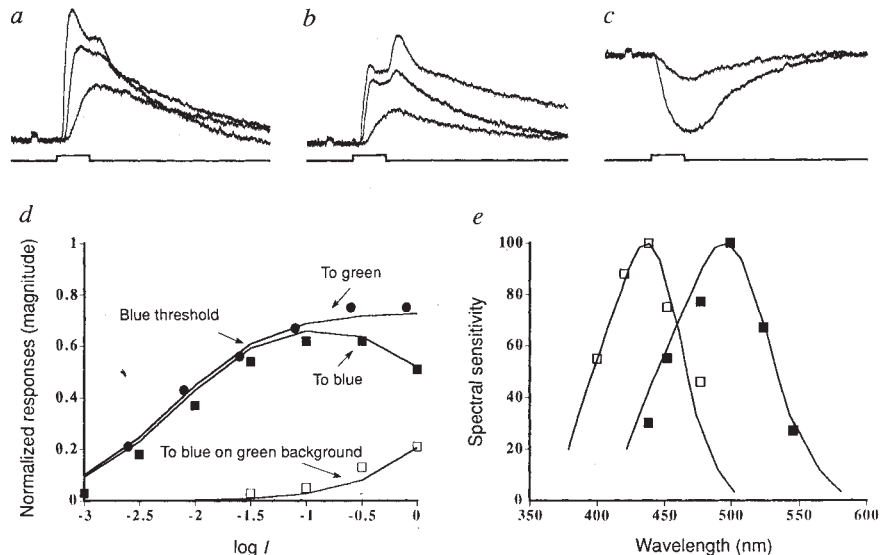


FIG. 1 The parietal eye of lizards is located in a small foramen between the parietal bones. The lens is immediately below a transparent cornea-like skin. Between the lens and the retina is the lumen into which project the photoreceptor outer segments. The retina has a single plexiform layer where processes of the photoreceptors (P) and ganglion cells (Ga) meet. Glial cells (Gi) span the retina.

FIG. 2 Intracellular responses and spectral sensitivity of a parietal eye photoreceptor in *Xantusia vigilis*. *a*, Depolarizing responses of a dark-adapted photoreceptor to green flashes (524 nm) of increasing intensity ($V_m = -50$ mV). At high intensities an 'on' transient is present. *b*, Responses of a dark-adapted photoreceptor to blue flashes (452 nm). With high intensity stimulation the initial intensity is cut short and a prominent depolarization at 'off' and slower decay are evident. *c*, Responses of a green (546 nm)-adapted photoreceptor to blue flashes. Responses are hyperpolarizing from the depolarized V_m of -36 mV. Flash duration, 1 s; calibration pulse in *a*, *b* and *c* was 1 mV. Glass or quartz micropipettes filled with 3M potassium acetate were advanced into the isolated parietal eye. The preparation was superfused with Earle's medium (Gibco), pH 7.4, equilibrated with 95% O_2 /5% CO_2 . *d*, Intensity-response functions from intracellular photoreceptor recordings reveal two antagonistic chromatic mechanisms. Normalized, dark-adapted responses to green (●) and to blue (■), and responses to blue flashes on green background (□). Note that these are response magnitudes for both the dark-adapted depolarizations and the green-adapted hyperpolarizations. In the dark at higher intensities, the amplitude of the green response increases faster than the blue, and the point of divergence correlates with the appearance of the 'off' transient seen in intracellular recordings. Thus we see a threshold level for the appearance of a second mechanism. Blue-induced hyperpolarizations in green-adapted photoreceptors were seen only at intensities above that threshold level. Solid lines represent response amplitudes predicted for a photoreceptor by a 3-state model with light-driven and thermal transitions. *e*, Action spectra for the green depolarizing mechanism (■) and the blue hyperpolarizing mechanism (□) under dark and green adaptation respectively. Solid lines represent Dartnall nomograms centred at 495 and 440 nm.



robinot positively identified photoreceptor cells as the source of these signals.

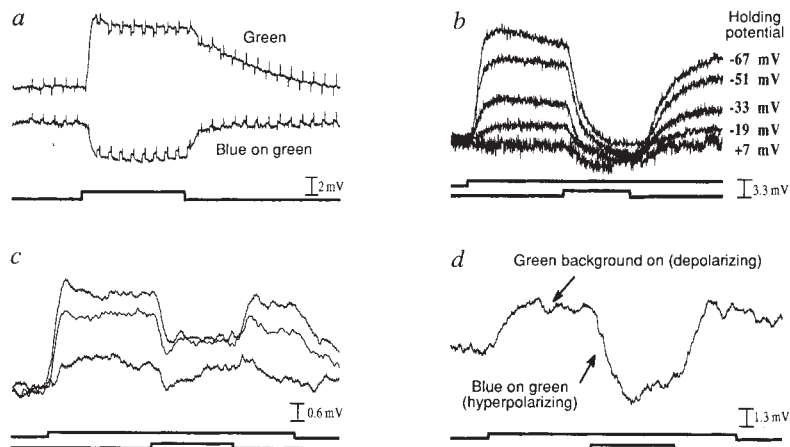
The opposing membrane potential changes are mediated by opposing conductance changes. In the dark-adapted condition, both blue and green stimuli decrease membrane resistance. As shown in Fig. 3a, a green stimulus produces a membrane resistance decrease of ~ 25 M Ω , resulting in membrane depolarization. However, a blue stimulus delivered to a green-adapted (and thereby depolarized) photoreceptor causes an increase in membrane resistance concurrent with a hyperpolarization back towards the dark resting membrane potential (Fig. 3a).

The wavelength-dependent conductance changes appear to result from the light-driven opening and closing of cation channels. Current-clamping the photoreceptor yields a reversal potential slightly above 0 mV for both the green and blue responses (Fig. 3b). This suggests a similar cationic dependency for both chromatic mechanisms. Moreover, replacing Na^+ with choline in the superfusion medium almost completely eliminates the intracellularly recorded photoresponses (Fig. 3c), indicating that both the green and blue mechanisms modulate conductances that are permeable at least to Na^+ . In short, green light induces a membrane depolarization mediated by an increased conductance to Na^+ , whereas a superimposed blue light decreases a similar membrane conductance, hyperpolarizing the membrane.

It is clear that the chromatic antagonism does not originate in synaptic interactions of two families of photoreceptors, but rather that the two spectral mechanisms reside in the same photoreceptor. We used enzymes to isolate single photoreceptors from the parietal eye of *U. stansburiana*, which were then used for recording: a green background depolarizes the cell and superimposed blue stimuli cause a relative hyperpolarization (Fig. 3d). Chromatic antagonism is seen in these photoreceptors even when potential synaptic input from neighbouring cells is eliminated.

To our knowledge, this is the first instance of a cilium-derived vertebrate photoreceptor with antagonistic spectral mechanisms which is therefore capable of depolarizing in response to light. Structural studies of the parietal eye have shown that these photoreceptors are derived from a ciliated cell^{1,10}, as are all other known vertebrate photoreceptors, yet years of intracellular

FIG. 3 *a*, Measurement of input impedance changes in *X. vigilis* parietal eye photoreceptor using intracellular micropipettes and bridge circuit. Measured input impedance in the dark was 120–250 M Ω . Upper trace: a 5-s green flash applied on a dark-adapted photoreceptor induces a depolarization accompanied by a resistance decrease (resistance typically decreased 15–20 M Ω). Lower trace: a 5-s blue flash superimposed on a green background induces a resistance increase accompanied by hyperpolarization. Resistance increases and decreases are of similar magnitude. *b*, Effects of current clamping on the antagonistic photoreceptor mechanisms. A photoreceptor was hyperpolarized or depolarized by injecting a steady current through the recording electrode. As the membrane potential is made more positive (resting potential was –50 mV) there is a similar decrease in amplitude of both the green-driven depolarization and the blue-driven hyperpolarization with a reversal potential ~ 0 mV. As the membrane potential was made more negative both depolarizing and hyperpolarizing responses increased in magnitude. However, their relative magnitudes were not altered. *c*, Effects of ionic replacement on the antagonistic photoreceptor mechanisms. Superimposed sequential photoresponses when Na⁺ concentration decreased as normal Earle's superfusion is replaced by choline-substituted Earle's. Amplitude of green and blue responses decreased similarly showing dependence on Na⁺. Amplitude recovered fully on return to normal Earle's. These results suggest that the light response is primarily due to the modulation of the influx of Na⁺ ions across the photoreceptor membrane. *d*, Representative whole-cell volt-



age recordings from an isolated parietal eye photoreceptor of *Uta stansburiana*. Green light elicited a depolarizing response whereas a superimposed blue flash resulted in a hyperpolarization. Chromatic antagonism originates within a single photoreceptor and is not the result of synaptic interactions. We repeated such recordings on photoreceptors of a second animal and confirmed the results shown. In *b*, *c* and *d*, the upper and lower stimulus traces are the applied green background (546 nm) and superimposed 2-s blue flash (452 nm), respectively.

recording have not revealed such a response in vertebrate retinal¹¹ or pineal photoreceptors^{12–16}.

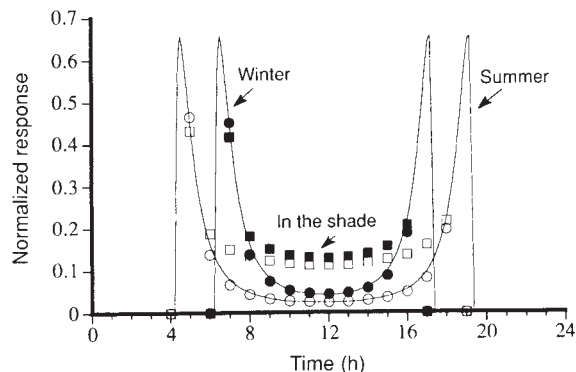
The basic biochemical mechanisms responsible for the chromatic antagonism remain unclear. A photoreceptor containing two photopigments with different absorption spectra, driving hyperpolarizing and depolarizing mechanisms, could produce the responses we have documented. The concentration of an internal messenger responsible for modulating membrane conductance could be regulated by the interactions of antagonistic G proteins coupled to the two photopigments. Alternatively, a bistable pigment, as proposed for photoreceptors in the frog frontal organ¹⁷, may be present. Bistable pigments have been noted in invertebrate photoreceptors such as barnacle¹⁸, *Limulus*¹⁹ and blowfly²⁰.

Although the functional significance of the parietal eye has been questioned¹, we can now incorporate electrophysiological data and speculate on a possible functional role. It is clear that chromatic antagonism modulates the ganglion cell discharge^{5,6}

and we find chromatic antagonism at the photoreceptor input to the ganglion cells. Based on the two spectral mechanisms described, the antagonistic nature of their interactions, and the disparity in their relative sensitivities, it is possible to attribute to this eye a potential role as a dawn/dusk detector. During the low light levels at dawn and dusk, only the more sensitive green mechanism is active and the photoreceptor depolarizes, almost reaching its saturation level. As the day proceeds and the light levels rise, then the antagonistic mechanism is activated and the photoreceptor hyperpolarizes back down close to its dark resting potential. The decreasing light levels throughout the afternoon eventually diminish the drive of the hyperpolarizing response, thereby generating a second depolarizing peak late in the day.

A mathematical model describing the intensity–response relations of the two antagonistic mechanisms of the photoreceptors was implemented. Spectral intensity values calculated for the visible spectrum²¹ in the natural environment of *X. vigilis* were

FIG. 4 Predicted steady-state membrane potential (solid lines) using a three-state model of a parietal eye photoreceptor over the course of a summer and a winter day in the natural environment of *X. vigilis*. Both summer and winter solar spectra produce enhancement of light–dark transitions. The steady-state photoreceptor responses predicted for direct (circles) and diffuse (squares) solar exposure are not sufficiently different to produce a depolarizing peak during normal activity as the animal shuttles between direct sunlight and shade to maintain body temperature (summer and winter represented by empty and filled symbols respectively). The model assumes that the photoreceptors can be shifted by light between G, a green-sensitive state, and B, a blue-sensitive state that drives the photoreceptor depolarization, or between B and X, a silent state. Thermal (not light-induced) transitions are allowed between G and B or between B and X. The system is described by a set of differential equations and solved for B under steady-state conditions such that $B_{ss} = M_{ss} / (I_g + I_b / m_{b_{ss}} + m_{g_{ss}})$, where B_{ss} is steady-state photoreceptor response, M_{ss} is maximum steady-state response (depolarization) of green stimulus on dark background, $m_{g_{ss}}$ is the intensity value of green on a dark background that elicits half maximum steady-state depolarization, and $m_{b_{ss}}$ is the intensity value of blue stimulus on a green background that elicits half maximum steady-state hyperpolarization, I_g is the green stimulus weighted by the green spectral sensitivity (Fig. 2e) and I_b is blue stimulus weighted by the blue spectral sensitivity



(Fig. 2e). B_{ss} predicts the intensity–response functions of the photoreceptors quite well (see solid lines in Fig. 2d). Values of spectral intensity for every hour in the day were calculated using SOLRAD²¹ and used as input to our model. Solid lines represent the estimated response for interpolated SOLRAD values using a cosine function.

used as input to the model. On a summer day, the model predicts a bimodal response pattern with dramatic depolarizing peaks in morning and evening hours (Fig. 4). The transitions between light and dark are emphasized. In the winter, when days are dimmer and shorter, the time interval between the response peaks is smaller and the response level at midday has increased. The model predictions further suggest that thermoregulatory behaviour involving shuttling between direct radiation and shade

will not generate the large depolarizing peaks seen at the beginning and end of a day. The intensity and spectral composition of the indirect skylight in the shade results in only a small increment in the response to light. These results suggest that the antagonistic nature of the photoresponses may constitute a mechanism to process diurnal light intensity and spectral composition to detect the beginning and end of the daily photophase. □

Received 31 March; accepted 12 May 1993.

1. Eakin, R. M. *The Third Eye* (Univ. California Press, Berkeley, 1973).
2. Engbretson, G. A. & Anderson, K. J. *Vis. Neurosci.* **5**, 395–404 (1990).
3. Jenison, G. & Nolte, J. *Brain Res.* **188**, 615–618 (1979).
4. Petit, A. Z. *Zellforsch.* **92**, 70–93 (1968).
5. Miller, W. H. & Wolbarsht, M. L. *Science* **135**, 316–317 (1962).
6. Dodt, E. & Scherer, E. *Vision Res.* **8**, 61–72 (1968).
7. Korf, H. W. & Wagner, U. *Cell Tiss. Res.* **219**, 567–583 (1981).
8. Engbretson, G. A., Reiner, A. & Brecha, N. J. *comp. Neurol.* **198**, 155–165 (1981).
9. Baylor, D. A. & Fuortes, M. G. F. *J. Physiol.* **207**, 77–92 (1970).
10. Ahmed, J. & Engbretson, G. A. *Vision Res.* (in the press).
11. Yau, K. W. & Baylor, D. A. *Rev. Neurosci.* **12**, 289–327 (1989).
12. Dodt, E. in *Handbook of Sensory Physiology* (ed. Jung, R.) 113–140 (Springer, Berlin, 1973).
13. Ekstrom, P. & Meissl, H. *Neuroscience* **25**, 1061–1070 (1988).

14. Hamasaki, D. I. & Dodt, E. *Pflugers Arch.* **313**, 19–29 (1969).
15. Pu, G. A. & Dowling, J. E. *J. Neurophysiol.* **46**, 1018–1038 (1981).
16. Tabata, M., Tamura, T. & Niwa, H. *Vision Res.* **15**, 737–740 (1975).
17. Eldred, W. D. & Nolte, J. *Vision Res.* **18**, 29–32 (1978).
18. Minke, B., Hochstein, S. & Hillman, P. J. *gen. Physiol.* **62**, 87–104 (1973).
19. Nolte, J. & Brown, J. E. *J. gen. Physiol.* **59**, 186–200 (1972).
20. Hamdorf, K. & Razmjoo, S. *Biophys. struct. Mech.* **5**, 137–161 (1979).
21. Porter, W. P. *SOLRAD and SINE: Clear Sky Direct and Diffuse Solar Radiation for Horizontal and Sloping Surfaces and Sinusoidal Air Temperature Variation over 24 Hours* (Univ. Wisconsin, Madison, 1989).

ACKNOWLEDGEMENTS. We thank R. B. Barlow and J. E. Lisman for reading early versions of the manuscript, the Sutter Instrument Co. at San Rafael for quartz micropipettes for intracellular recording, and especially D. Sze for help in developing the isolated photoreceptor preparation.

Inhibition of axonal growth by SNAP-25 antisense oligonucleotides *in vitro* and *in vivo*

Astrid Osen-Sand*, Marina Catsicas†, Julie K. Staple*, Kenneth A. Jones*, Guidon Ayala*, Jonathan Knowles*, Gabriele Grenningloh* & Stefan Catsicas*‡

* Glaxo Institute for Molecular Biology S.A. 14, chemin des Aulx, Case Postale 674, 1228 Plan-les-Ouates, Geneva, Switzerland

† Institute of Anatomy, 9 rue du Bugnon, 1005 Lausanne, Switzerland

AXONAL elongation and the transformation of growth cones to synaptic terminals are major steps of brain development and the molecular mechanisms involved form the basis of the correct wiring of the nervous system. The same mechanisms may also contribute to the remodelling of nerve terminals that occurs in the adult brain, as a morphological substrate to memory and learning¹. We have investigated the function of the nerve terminal protein SNAP-25 (ref. 2) during development. We report here that SNAP-25 is expressed in axonal growth cones during late stages of elongation and that selective inhibition of SNAP-25 expression prevents neurite elongation by rat cortical neurons and PC-12 cells *in vitro* and by amacrine cells of the developing chick retina *in vivo*. These results demonstrate that SNAP-25 plays a key role in axonal growth. They also suggest that high levels of SNAP-25 expression in specific areas of the adult brain² may contribute to nerve terminal plasticity.

SNAP-25 (synaptosomal-associated protein 25, M_r 25K) is a neuron-specific protein, differentially expressed in the nervous system^{2,3} and translocated to terminals by fast axonal transport^{4,5}. SNAP-25 expression is strongly induced during synapse formation^{6,7} but low levels are detectable earlier during development^{8,9}. SNAP-25 may also be involved in axonal growth and differentiation¹⁰ and here we test this hypothesis. We first analysed SNAP-25 expression in newborn rat cortical cells in culture and compared it to that of cytoskeletal and synaptic proteins (manuscript in preparation). We found that SNAP-25 is expressed in the growth cone and distal segment of

developing axon-like neurites, after their initial outgrowth but before the formation of synaptic boutons (Fig. 1a–c). To block SNAP-25 expression we added to the culture medium modified antisense phosphorothioate oligonucleotides (PONs) complementary to position 1–20 or 533–552 (relative to the translation initiation site) of the murine SNAP-25 coding region². Controls included the two sense and one unrelated PONs. PONs were added on the first day *in vitro*, and after one week SNAP-25 expression was blocked with both antisense PONs to levels undetectable with immunohistochemistry. Inhibition of SNAP-25 expression reached 94% and was dose-dependent (Fig. 1d,e). The expression of control proteins was not affected but the level of synaptophysin was decreased by more than 60% (Fig. 1d,e). In addition, qualitative analysis of the cultures showed a strong reduction of the network of thin, axon-like neurites following the antisense treatment (not shown).

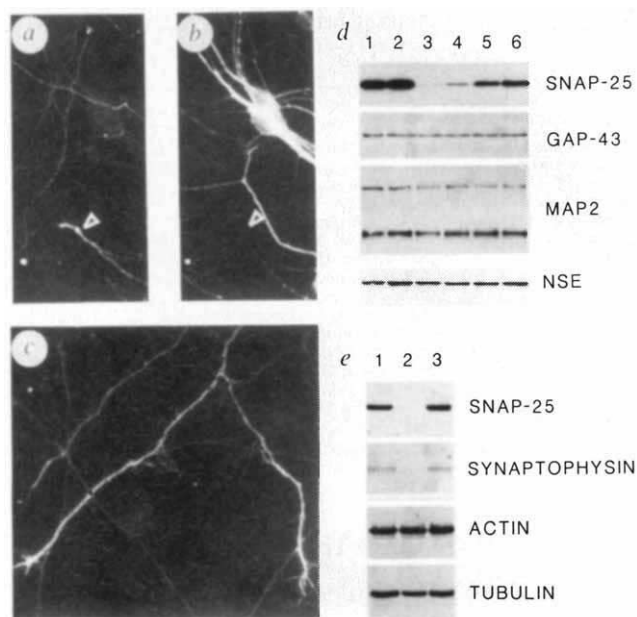
Synaptophysin is a marker of differentiated synapses¹¹, so the effects on its level of expression suggest that inhibition of axonal growth reduced synapse formation. But a precise quantification of axonal growth with image-analysis (see below) was difficult because axons and dendrites intermingled. Also, the role of SNAP-25 in axonal elongation may be underestimated in these cultures because several days of PONs treatment are necessary to inhibit SNAP-25 expression and by that time neurons have already sent axons and start to form synaptic contacts. For a more quantitative assay of the SNAP-25 role in neurite elongation, we used PC12 rat pheochromocytoma cells¹², because: (1) all of the cells express SNAP-25¹³; (2) all the nerve growth factor (NGF)-induced neurites are SNAP-25 positive¹³; and (3) the cells can be incubated with PONs before the induction of neurite growth with NGF. The same PONs used for primary neurons were used for the PC12 cells. Both SNAP-25 antisense PONs but not control PONs specifically reduced SNAP-25 expression by more than 75% and this effect was dose-dependent (Fig. 2a). To assay neurite growth, the cells were treated with PONs for few days and were then induced with NGF, in the presence of antisense and control PONs, to determine their ability to extend neurites. Controls included NGF-induced and non NGF-induced cultures that were not treated with PONs. In all cultures, the cells started to extend neurites 3 days after NGF induction. After 4 additional days, the cultures were fixed, immunostained and neurite growth was analysed quantitatively. The two SNAP-25 antisense PONs inhibited neurite extension to levels found in PC12 cells not induced with NGF, whereas

‡ To whom correspondence should be addressed.

control PONs had no significant effects (Fig. 2*b*). The number of neurites extended by the cells was not affected by PON treatment. These results show that SNAP-25 has a major role in neurite elongation and is apparently not involved in initial

outgrowth. In addition, inhibition of SNAP-25 expression had no effect on the elaboration of dibutyryl cyclic AMP-induced spikes (not shown). The formation of these short processes occurs through a different mechanism than the NGF-induced

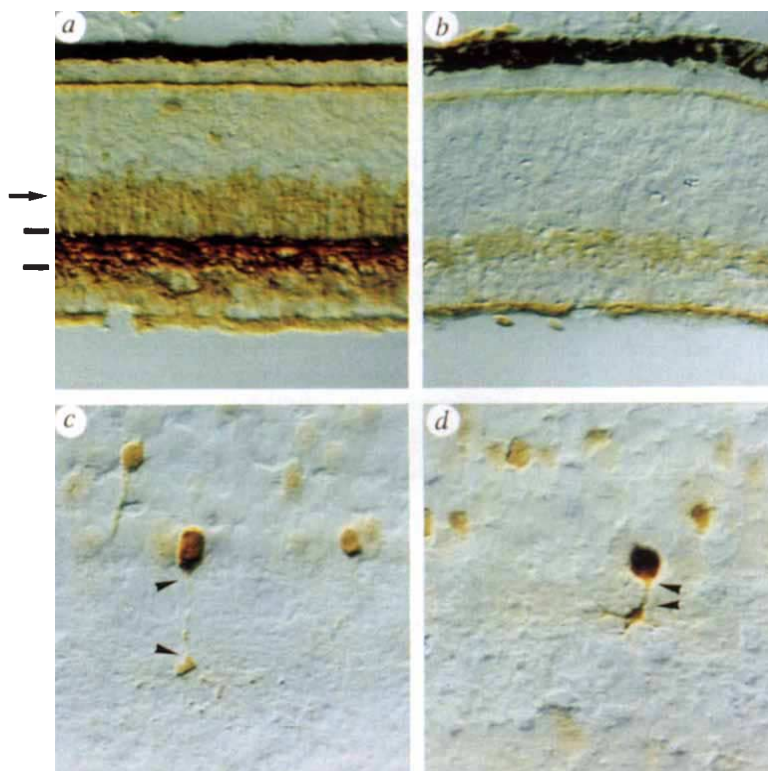
FIG. 1 *a–c*, SNAP-25 expression by cortical neurons in culture. *a*, *b*, Double labelling after 5 days in culture with SNAP-25 (*a*) and MAP2 (*b*) antibodies. MAP2 is localized in cell bodies and thick processes, whereas SNAP-25 is found selectively in thin axon-like neurites. The open arrowheads indicate the position of a SNAP-25-positive growth cone. *c*, At high power, SNAP-25 immunoreactivity appears to reach the highest intensity in the growth cone. *d,e*, Effects of oligonucleotide treatment on primary neurons. *d*, Dose-response and specificity of the antisense treatment. Western blots of protein extracts from cultures incubated for 7 days with 2 μ M control (sense) PONs (lane 1), without PONs (lane 2), or with 2, 1, 0.5 and 0.25 μ M antisense PONs (oligo 1-20), lanes 3, 4, 5, and 6, respectively. *e*, Effects of SNAP-25 expression inhibition on synaptophysin expression. Western blots of protein extracts incubated 7 days without PONs or with 2 μ M antisense (oligo 1-20) and control (sense) PONs, lanes 1, 2 and 3, respectively. METHODS. Neurons from primary visual cortex of postnatal day 1 rats were plated on astrocyte monolayers²². When treated with PONs, the cells were plated in medium containing 2 μ M PON. Then, PONs (0.25 to 2 μ M final concentration each time) were added every day. For immunocytochemistry, cultures were fixed with 4% paraformaldehyde at 1-day intervals for a period of 10 days and stained with appropriate antibodies (see below). Protein expression was visualized with indirect immunofluorescence. For western blotting, total protein was extracted at 2-day intervals with standard procedures, separated on SDS-PAGE (10%), transferred to nitrocellulose and incubated with antibodies against SNAP-25², GAP-43 (Boehringer Mannheim), MAP2 (Sigma), neuron-specific enolase (NSE; Sera-Lab), synaptophysin (Boehringer Mannheim), tubulin (Boehringer Mannheim) and actin (Boehringer Mannheim). Binding of the antibodies was visualized with the ECL Western Blotting Kit (Amersham), quantified with an image analyser



(Vidas, Kontron) and normalized relative to total protein levels. The same blot was used for three or four different antibodies.

FIG. 3 Inhibition of axonal elongation *in vivo*. Transverse sections of the retinae of one embryo that received two injections of 10 μ M PON and was fixed at day 13. *a,b*, SNAP-25 immunostaining of the control (*a*) and antisense-injected (*b*) retina. Note the decrease of immunostaining in the IPL (bars) and in the cell bodies of the amacrine cells (arrow); and the reduction in the thickness of the IPL. *c,d*, Calbindin-positive amacrine cells of the control (*c*) and antisense-injected (*d*) retina. Note the difference in the length of the main process (arrowheads). The effects on the parameters analysed in this embryo were as follows. Densitometric analysis of SNAP-25, tubulin and synaptophysin expression in the IPL: 71%, 2% and 33% reduction, respectively; IPL thickness: 52% reduction; length of processes: 49% reduction.

METHODS. Eggs were opened and PONs injected in the vitreous body with a Hamilton syringe²³. PONs final concentration in the eye after each injection ranged from about 1 to 10 μ M. The embryos received two injections into one eye (at days 9 and 11) and were killed at day 12 or 13. Eyes were removed and fixed with 4% paraformaldehyde. Frozen sections from both eyes of each embryo were mounted on the same slide and processed for immunohistochemistry. Serial sections were stained with antibodies against SNAP-25², tubulin (Boehringer Mannheim), calbindin (Swiss Antibodies) or synaptophysin (Boehringer Mannheim) and immunoreactivity was viewed with indirect immunofluorescence or with the avidin-biotin peroxidase method (Vector). The effects of the injections on the thickness of the retinal layers and on the intensity of immunoreactivity with the antibodies used were measured on avidin-biotin reacted sections with computer-assisted image analysis (Vidas, Kontron). For all measurements we selected corresponding areas of the central retina in the experimental and control eyes. The thickness of the retinal layers was measured on the tubulin-immunostained sections viewed with Nomarski optics. For the optical



density analysis in the IPL we measured and averaged the mean grey value in 3 strips about 60 μ m wide for each retina.

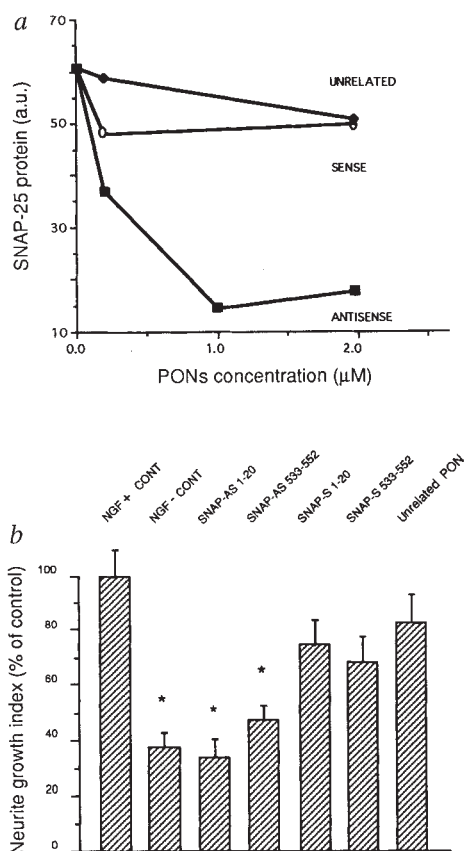


FIG. 2 **a**, Dose-response curve of SNAP-25 expression inhibition in PC12 cells with antisense (oligo 1–20) and control (sense and unrelated) PONs. a.u., Normalized mean grey value of the signal in western blots in arbitrary units. **b**, Inhibition of neurite extension by PC12 cells. SNAP-25 antisense oligonucleotides prevent NGF-induced neurite elongation, whereas control oligonucleotides do not. To display all the data on the same figure, the index of neurite growth is shown as percentage of the NGF-stimulated control for all culture conditions tested. The asterisks indicate statistically significant differences from the NGF-stimulated controls with ($P < 0.05$) and without ($P < 0.01$) PONs, as calculated with the original data (t -test). Error bars indicate standard error of the mean. NGF+CONT, NGF-stimulated control; NGF-CONT, non-NGF-stimulated control.

METHODS. PC12 cells were maintained in culture as described¹². To establish the dose-response the cells were plated in medium containing NGF (25 ng ml⁻¹) and 10 μM PON. Then, PONs (0.2 to 2 μM final concentration each time) were added every day. Total proteins were extracted and antibody binding visualized and quantified as described for Fig. 1. For the neurite growth assays, cells were plated on collagen-coated 24-well Costar plates with medium containing 10 μM PONs and 2 μM PONs were added every two days. The sequence of the unrelated PON was: 5'-ATGCTGTGCTGTATGAGAAG-3'. NGF (25 ng ml⁻¹) was added to the cultures for the first time on day two and then every two days. Each plate contained wells with different culture conditions in duplicate (antisense, sense, unrelated or no PONs) and had NGF-stimulated wells that were not treated with PONs as internal control. The cultures were fixed 7 days after NGF induction with 4% paraformaldehyde and immunostained to check the inhibition of SNAP-25 expression. The plates were then dried and the cells analysed under phase-contrast using a semi-automatic computer-assisted image analyser (Vidas, Kontron). All the images were digitized under constant conditions and the boundaries of the cell bodies were determined on the digitized image. The index of neurite growth, defined as the ratio between the surface area of neurites and the surface area of cell bodies, was then calculated automatically. The number of measurements in each well ranged from 43 to 74. The surface area of the cell bodies was not affected by PON treatment. PONs were synthesized according to Applied Biosystems, using their DNA synthesizer 394-8.

neurite elongation¹⁴ and this result is consistent with our interpretation.

To study the role of SNAP-25 *in vivo*, we used the antisense PON complementary to position 533–552, a region identical in the murine and chick transcripts⁶, for intraocular injections in chick embryos. Injections of antisense or sense PONs were timed to inhibit SNAP-25 synthesis in the retina over 4 days, at the time when process growth leads to the formation of the inner plexiform layer (IPL). The IPL is the region of the retina where the density of growing axons and the intensity of SNAP-25 expression is highest⁶ and the thickness of the IPL is directly related to process growth. Retinae injected with the sense PON ($n=4$) were similar (less than 15% difference) to control non-injected retinae for all the parameters measured (see below). In contrast, following antisense injections ($n=5$), the intensity of SNAP-25 immunoreactivity in the remaining IPL was reduced by 49–77% and the thickness of the IPL was reduced by 52–87% (Fig. 3). The reduction of the IPL thickness correlated with the reduction of SNAP-25 expression. Thus, inhibition of SNAP-25 expression in the retina reduces axonal growth. The effects on the intensity of tubulin immunoreactivity in the remaining IPL were not significant and varied from a 2% decrease to a 10% increase. To study the effects of the inhibition of SNAP-25 expression on the morphology of identified neurons, the retinae were also stained with a monoclonal antibody against calbindin¹⁵. This allowed us to identify a subtype of calbindin-positive amacrine cell¹⁶ located near the edge of the IPL. These cells have a single process that crosses the entire IPL, forms a characteristic bulge and then branches. The number and distribution of these cells was not affected by the PONs injections but their processes were shorter in the antisense-injected retinae (Fig. 3). In these retinae, there was a mean reduction of 42–73% in the distance between the cell body and the branching point ($n=23–45$; $p < 0.01$, t -test), consistent with the reduction in thickness of the IPL. Terminal branches also seemed shorter but were not measured. To estimate further the stage of differentiation of the terminals, we measured the intensity of synaptophysin immunoreactivity¹⁷ and found a reduction ranging from 33 to 56% in the remaining IPL of the antisense-injected retinae. These findings are identical to our results with primary neurons *in vitro* and demonstrate that SNAP-25 plays an essential role in axonal elongation during development. SNAP-25 is also found at high levels in specific areas of the adult brain, including regions of synaptic plasticity^{2,8}. Although the precise function of the protein in adults is still unknown^{18,19}, its distribution and our data suggest that SNAP-25 could contribute to axonal terminal remodelling, a process involved in cognitive functions and pathologies of the brain.

Inhibition of SNAP-25 expression *in vivo* affected the length and terminal differentiation of specific processes, but not their direction of growth nor the general morphology of the neurons. This selectivity shows that it is possible to use antisense oligonucleotides to dissect the molecular mechanisms of complex cellular behaviours *in vivo*. Finally, Söllner *et al.*²⁰ have reported that SNAP-25 is an essential member of the cell fusion apparatus in adult neurons. Taken together, our data suggest that the role of SNAP-25 in axonal elongation is to promote new membrane insertion by mediating fusion of the vesicles that constitute the plasmalemmal precursor pool in growth cones²¹. In adults, the same mechanism could underlie plasticity or membrane turnover and repair in nerve terminals. □

Received 4 December 1992; accepted 26 May 1993.

- Wallace, S. W. *et al.* in *Long-Term Potentiation: a Debate of Current Issues* (eds Baudry, M. & Davis, J. L.) 189–232 (MIT press, Cambridge, MA 1991).
- Oyler, G. A. *et al.* *J. Cell Biol.* **109**, 3039–3052 (1989).
- Catsicas, S. *et al.* *J. Neurosci. Res.* **33**, 1–9 (1992).
- Loewy, A. *et al.* *J. Neurosci.* **11**, 3412–3421 (1991).
- Hess, D. T. *et al.* *J. Neurosci.* **12**, 4634–4641 (1992).
- Catsicas, S. *et al.* *Proc. nat. Acad. Sci. U.S.A.* **88**, 785–789 (1991).
- Oyler, G. A. *et al.* *Proc. nat. Acad. Sci. U.S.A.* **88**, 5247–5251 (1991).

8. Oyler, G. A. et al. *Dev. Brain Res.* **65**, 133-146 (1992).
9. Bark, C. & Wilson, M. C. *Abstr. Soc. Neurosci.* **17**, 531 (1991).
10. Alvarez-Bolado, G. et al. *Abstr. Soc. Neurosci.* **17**, 215 (1991).
11. Sudhof, T. C. et al. *Science* **238**, 1142-1144 (1987).
12. Greene, L. A. & Tischler, A. S. *Proc. nat. Acad. Sci. U.S.A.* **73**, 2424-2428 (1976).
13. Sanna, P. P., Bloom, F. E. & Wilson, M. C. *Dev. Brain Res.* **59**, 104-108 (1991).
14. Gunning, P. W. et al. *J. Cell Biol.* **89**, 240-245 (1981).
15. Celio, M. R. *Neuroscience* **35**, 375-475 (1990).
16. Ellis, J. H., Richards, D. E. & Rogers, J. H. *Cell Tissue Res.* **264**, 197-208 (1991).
17. Masliah, E. et al. *J. Histochem. Cytochem.* **38**, 837-844 (1990).
18. Geddes, J. W. et al. *Neuroscience* **38**, 515-525 (1990).
19. Hess, E. J. et al. *J. Neurosci.* **12**, 2865-2874 (1992).
20. Söllner, T. et al. *Nature* **362**, 318-324 (1993).
21. Pfenninger, K. H. et al. in *The Growth Cone* (eds Letourneau, P. C., Kater, S. B. & Macagno, E. R.) 111-123 (Raven, New York, 1991).
22. Baughman, R. W. et al. in *Culturing nerve cells* (eds Banker, G. & Goslin, K.) 227-249 (MIT Press, Cambridge, MA, 1991).
23. Catsicas, S. & Clarke, P. G. H. *J. comp. Neurol.* **262**, 512-522 (1987).

ACKNOWLEDGEMENTS: We thank P. Schwab for computer programs; D. Talabot for help with image analysis; W. Lukas for help with cortical cultures; C. Hebert for help with photography; M. C. Wilson for SNAP-25 antisera, sharing unpublished observations and for critical reading of the manuscript; P. G. H. Clarke, C. S. Goodman, S. Ketchum, E. Kawashima, J. C. Martinou and R. Sadoul for many suggestions at various stages of this work. M. C. was supported by a grant from the Swiss National Science Foundation to P. G. H. Clarke.

A cluster of three GABA_A receptor subunit genes is deleted in a neurological mutant of the mouse *p* locus

Yoshimichi Nakatsu*, Rachel F. Tyndale†‡, Timothy M. DeLorey†, Donna Durham-Pierre*, John M. Gardner*, Heather J. McDanel*, Q. Nguyen†, Joseph Wagstaff§||, Marc Lalande§||¶, James M. Sikela#, Richard W. Olsen†, Allan J. Tobin‡ & Murray H. Brilliant*☆

* The Institute for Cancer Research, Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, Pennsylvania 19111, USA

† Department of Pharmacology, School of Medicine, University of California, Los Angeles, California 90024, USA

‡ Department of Biology, University of California, Los Angeles, California 90024, USA

§ Genetics Division, Children's Hospital,

|| Department of Pediatrics, Harvard Medical School, and

¶ Howard Hughes Medical Institute, Boston, Massachusetts 02115, USA

Department of Pharmacology, University of Colorado Health Science Center, Denver, Colorado 80262, USA

THE mouse pink-eyed cleft-palate (*p^{cp}*) mutation is characterized by hypopigmentation associated with cleft palate, neurological disorders and runting^{1,2}. Most *p^{cp}* homozygotes are born with cleft palate and die shortly after birth, presumably as a result of feeding problems³. A few exceptional *p^{cp}* mutants live beyond this stage but display tremor and jerky gait². We report here that the genes encoding the γ -aminobutyric acid type A (GABA_A) receptor subunits $\alpha 5$ (originally described as $\alpha 4$; ref. 4), $\beta 3$ and $\gamma 3$ are disrupted by a deletion in *p^{cp}* mice. We also show that the $\alpha 5$ and $\gamma 3$ genes are located between the *p* and $\beta 3$ genes on mouse chromosome 7. The *p^{cp}* deletion leads to alterations of binding properties of the GABA_A receptors in the brain, providing an *in vivo* model system for studying GABA_A receptor function. The human homologue of the region deleted in *p^{cp}* mice is associated with Angelman syndrome⁵⁻⁹. Thus, *p^{cp}* mice may be useful in defining the region containing the gene(s) for this syndrome.

GABA_A receptors, which mediate the majority of inhibitory synapses in brain, are ligand-gated ion channels composed of subunits encoded by a gene family of at least 15 members¹⁰.

GABA_A receptor subunit genes, $\beta 3$ and $\alpha 5$, have been mapped to human chromosome 15q11-13 (refs 5, 9). Because this region is homologous to mouse chromosome 7 near the *p* locus¹¹⁻¹⁵, we considered that mutation of the $\beta 3$ gene and/or the $\alpha 5$ gene might lead to neurological deficiencies of *p^{cp}* mice. We investigated this idea by Southern blot analysis (Fig. 1). Genomic sequences were detected by a 5' portion of $\beta 3$ cDNA in *p^d* (control) but not in *p^{cp}*, whereas the 3' portion of $\beta 3$ cDNA hybridized to both *p^d* and *p^{cp}* DNA. The entire $\alpha 5$ gene was absent in homozygous *p^{cp}* mice. These results indicate that the 5', but not the 3' portion of the $\beta 3$ gene and the entire $\alpha 5$ gene are deleted in *p^{cp}* mice. Gene expression studies showed that other GABA_A receptor subunit genes ($\alpha 1$, $\alpha 2$, $\alpha 3$, $\alpha 4$, $\alpha 6$, $\beta 1$, $\beta 2$, $\gamma 1$, $\gamma 2$ and δ) were expressed in *p^{cp}* brain, with the exception of the $\gamma 3$ subunit (R.F.T., unpublished result). Southern blot analysis revealed a deletion of the $\gamma 3$ gene in *p^{cp}* mice (Fig. 1), indicating that the $\gamma 3$ gene is closely linked to the $\alpha 5$ and $\beta 3$ genes near the *p* locus on mouse chromosome 7. Interestingly, these three clustered genes show similar temporal expression patterns in development¹⁶. Clustering of GABA_A receptor subunit genes ($\alpha 2$, $\beta 1$ and $\gamma 1$ genes on human chromosome 4) has also been proposed¹⁷.

We have isolated and characterized a complementary DNA representing the mouse *p* gene and shown that most of the *p* gene is deleted in *p^{cp}* mice¹⁸. Southern blot analysis reveals that the presumptive exon 1 is present in *p^{cp}* (Fig. 1). However, the sequence corresponding to the 3' end of *p* cDNA is completely missing in *p^{cp}* (Fig. 1). These results, in combination with the GABA_A receptor gene studies, indicate that the *p^{cp}* deletion extends from the first intron of the *p* gene through the $\gamma 3$ and $\alpha 5$ genes, and includes the 5' part of the $\beta 3$ gene (Fig. 2). Thus, the *p* gene and the $\beta 3$ gene are in the same transcriptional orientation, with the $\gamma 3$ and $\alpha 5$ genes located between them.

Our results suggested that a novel configuration of genetic information was created in *p^{cp}*, in which the 5' end of the *p* gene was fused to the 3' part of the $\beta 3$ gene, therefore creating a chimaeric *p*- $\beta 3$ gene driven by the *p* gene promoter. As the *p*

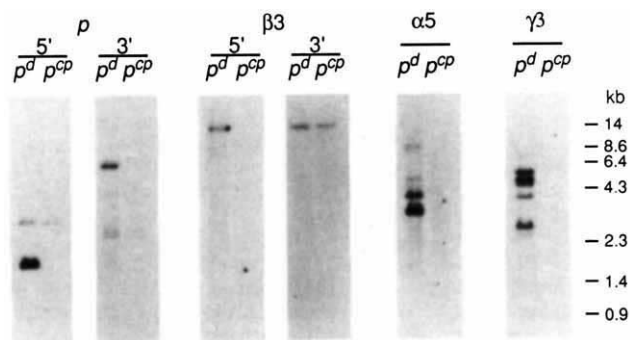


FIG. 1 Southern blot analyses of *p^{cp}/p^{cp}* and *p^d/p^d* genomic DNA. The *p^{cp}* mutation is maintained by intercrossing *p^{cp}/p^d* heterozygotes. *p^d/p^d* mice are used here as a control and are characterized by intermediate pigmentation, with no abnormal neurological phenotype^{1,2}. DNAs from homozygous *p^{cp}* and *p^d* mice were digested with *Hind*III, separated by electrophoresis on a 0.7% agarose gel, blotted to HybondN+ (Amersham) membranes and hybridized to the indicated probe sequences: 5' *p* represents a 424-bp *Eco*RI fragment from the 5' end of *p* cDNA clone MC2701 (ref. 18). This probe detects three *Hind*III fragments in homozygous wild type and *p^d* DNAs. The 3.2-kb *Hind*III fragment contains the putative first exon that corresponds to the region from nucleotide number 1 to 111 of *p* cDNA clone MC2701; 3' *p* represents a 363-bp *Sst*I fragment from the 3' end of MC2701; 5' $\beta 3$ indicates 170-bp *Sma*I fragment from mouse $\beta 3$ cDNA (J.W. and M.L., unpublished); 3' $\beta 3$, a 230-bp *Eco*RI/*Hpa*I fragment from mouse $\beta 3$ cDNA; $\alpha 5$, rat $\alpha 5$ cDNA⁴; $\gamma 3$, mouse $\gamma 3$ cDNA²⁷. Hybridization was with $\sim 2-4 \times 10^6$ c.p.m. ml⁻¹ probe in phosphate-buffered 7% SDS solution at 62.5 °C. Blots were washed with 0.2 × SSC/0.1% SDS (final) at 62.5 °C before exposure to X-ray film.

☆ To whom correspondence should be addressed.

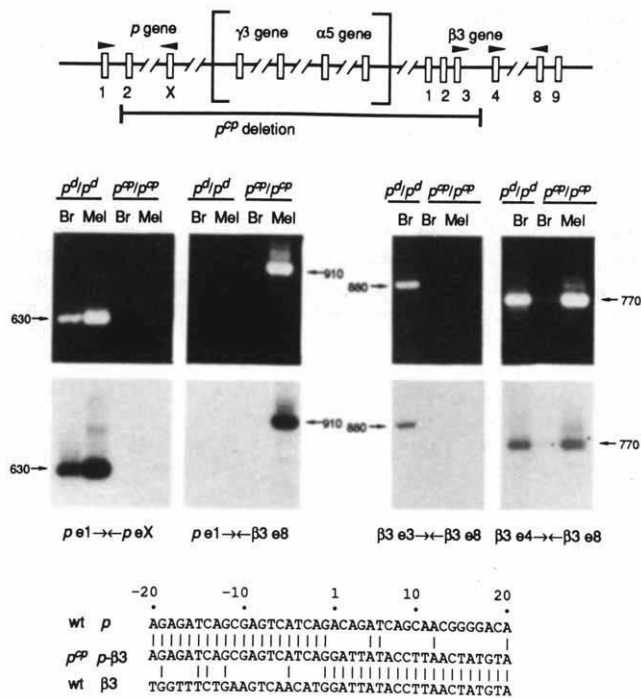


FIG. 2 Analysis of transcripts of *p* and $\beta 3$ in homozygous *p*^{cp} and *p*^d mice. The predicted organization of mouse *p*; $\gamma 3$, $\alpha 5$ (relative order unknown); and $\beta 3$ genes is shown at the top, indicating the region deleted in the *p*^{cp} genome. Exons of $\beta 3$ were deduced from the exon/intron structure of the gene for GABA_A receptor subunit $\beta 1$ (ref. 28). The location and orientation of primers used in RT-PCR experiments are shown above the appropriate exons. Nucleotide positions of the primers in their respective cDNAs^{18,29} are listed: *p* e1, 5'-CTTACACCAGGGTTG-TGCTCCATCC-3' (*p* cDNA nucleotides 21–45); *p* eX, 5'-ACCGCACACAGC-ACCCAAGCTT-3' (*p* cDNA antisense nucleotides 626–647); $\beta 3$ e3, 5'-AGTCTGCGTGGGGATGAACATCGAC-3' (rat $\beta 3$ cDNA nucleotides 250–274); $\beta 3$ e4, 5'-CGCCTACTCTGGGATCCCTCTCAACCTCACG-3' (rat $\beta 3$ cDNA nucleotides 361–391); $\beta 3$ e8, 5'-TCGATCATTCTTGGCCTTGGCTGT-3' (rat $\beta 3$ cDNA antisense nucleotides 1,106–1,129). The middle panels depict RT-PCR analysis of brain (br) and cultured melanocyte (mel; D.D.-P., unpublished) transcripts from homozygous *p*^{cp} and *p*^d mice assayed by the indicated primer pairs. METHODS. RT-PCR was performed as described³⁰ with the modification that cDNAs were synthesized from 2.5 μ g total RNA by MMLV reverse transcriptase (Gibco/BRL) with 100 pmol random hexamers (Pharmacia/LKB). For PCR amplification, 100 ng cDNA, 100 pmol of each primer and 2 units of Vent polymerase (NEB) were used in a 100- μ l reaction. PCR was done under the following conditions: denaturation at 95 °C for 1 min; annealing at 56 °C for 1 min; polymerization at 72 °C for 1 min; repeated 35 times. PCR products were separated on a 2% agarose gel. To analyse the PCR products, an oligomer probe (*p* cDNA antisense nucleotide 56–91) was used for *p* and *p*- $\beta 3$ blots and mouse $\beta 3$ cDNA for $\beta 3$ blot analysis. The chimaeric *p*- $\beta 3$ PCR products were also detected by mouse $\beta 3$ cDNA probe (data not shown). The 910-bp chimaeric cDNA was specifically amplified using *p*^{cp} melanocyte RNA and was isolated, cloned and sequenced. It is compared with the wild-type sequence of the *p* gene (top line; negative numbering, exon 1; positive numbering, exon 2) and rat $\beta 3$ (bottom line; exon 3, negative numbering; exon 4, positive numbering). Nucleotide number 1 corresponds to nucleotide 112 of *p* cDNA and to nucleotide 311 of rat $\beta 3$ cDNA.

gene is primarily expressed in melanocytes and only slightly in brain¹⁸, it is conceivable that this chimaeric gene is expressed in *p*^{cp} melanocytes and brain. Indeed, we were able to detect this predicted chimaeric transcript by reverse transcriptase-polymerase chain reaction (RT-PCR) (Fig. 2). A combination of *p* and $\beta 3$ primers specifically amplified a cDNA of 910 base pairs (bp) using RNA from *p*^{cp} melanocytes (a small but detectable amount of the cDNA was reproducibly amplified from *p*^{cp} brain RNA); RNA from control littermates failed to produce an amplification

product using primers from these two different genes. Two observations from these experiments suggest that the second breakpoint of the *p*^{cp} deletion is within intron 3 of $\beta 3$. First, sequence analysis of the PCR product showed that the amplified cDNA is composed of both *p* and $\beta 3$ sequences, with the fusion of exon 1 of *p* to exon 4 of $\beta 3$ (Fig. 2 legend). Second, $\beta 3$ e3 and e8 primers failed to amplify any cDNA from *p*^{cp} melanocytes, whereas $\beta 3$ e4 and e8 primers did (Fig. 2). The deduced product of this chimaeric gene is a truncated $\beta 3$ subunit, missing 85 amino acids from the amino terminus. As this chimaeric gene is transcribed in *p*^{cp} brain, albeit at low levels, it might cause *p*^{cp} phenotypes. However, normal phenotypes are observed in heterozygotes.

The pharmacological consequences of deficits in these GABA_A receptor subunits were analysed by ligand binding (Fig. 3). Consistent with the $\alpha 5$, $\beta 3$ and $\gamma 3$ subunit messenger RNAs being major GABA_A gene products in prenatal rodent brain^{16,19}, their absence in prenatal *p*^{cp} mice produces a significant decrease (~80%) of GABA_A receptor content, as determined by levels of the benzodiazepine antagonist [³H]Ro15-1788 binding (Fig. 3a). Co-expression of the $\alpha 5$ subunit with any β and the $\gamma 2$ subunit in cultured cells has been reported to produce receptors with much lower affinity (100-fold) for the benzodiazepine zolpidem as compared to receptors containing other α -subunits²⁰. In agreement with this observation, *p*^{cp}/*p*^{cp} hippocampus lacked low-affinity binding to zolpidem (Fig. 3b). An intermediate effect was noted in heterozygous *p*^d/*p*^{cp} mice (data not shown).

The importance of GABA_A receptors is underscored by a point mutation of the $\alpha 6$ subunit gene that drastically alters the ligand-binding properties of GABA_A receptors, possibly causing impaired motor control in an alcohol-non-tolerant rat line²¹. Significantly, $\alpha 5$, $\beta 3$ and $\gamma 3$ subunits are predominantly expressed in fetal and neonatal rodent brain but diminished in

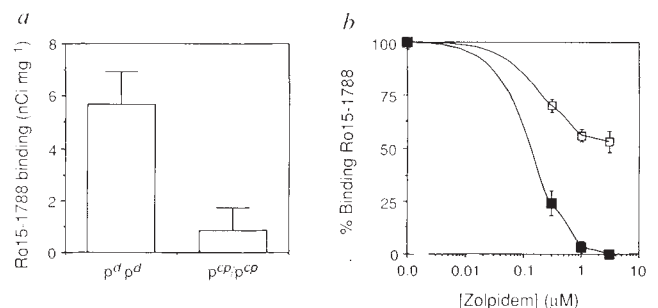


FIG. 3 a, Specific binding of the benzodiazepine antagonist [³H]Ro15-1788 in hippocampus of prenatal (18-day-old fetus) *p*^d and *p*^{cp} homozygous mice. b, Inhibition of [³H]Ro15-1788 binding by zolpidem in hippocampus of neonatal *p*^d (□) and *p*^{cp} (■) homozygous mice (at this stage, an overall decrease of 44% of specific binding was observed). Data derived from film autoradiograms of sagittal mouse brain sections labelled with [³H]Ro15-1788. No gross morphological differences were observed between mutant and control brain sections. A more detailed analysis is pending and we² have noted overall size differences between *p*^{cp} homozygous mice and controls.

METHODS. Slide-mounted sagittal sections of both *p*^d and *p*^{cp} homozygous mice were preincubated on ice for 30 min in 170 mM Tris-HCl pH 7.4. Sections were incubated for 60 min on ice in Tris buffer with 1 nM [³H]Ro15-1788 in combination with either buffer alone or one of the benzodiazepines clonazepam (1 μ M) or zolpidem (0.3, 1 or 3 μ M). Binding was terminated by washing twice for 1 min each with ice-cold buffer. Sections were dried by a steady stream of air passed through a desiccant. Sections were exposed to UltraFilm (Pharmacia/LKB) for 10 d. The film was developed and autoradiograms analysed by computer to determine radioligand binding (for details, see ref. 31). The overall binding of the radioligand is reduced in the *p*^{cp} mice relative to the *p*^d mice, therefore data are normalized for comparison. Specific binding is defined as total [³H]Ro15-1788 binding minus the nonspecific binding of [³H]Ro15-1788, measured in the presence of 1 μ M clonazepam.

most areas of adult brain^{16,19}, suggesting that they may contribute to cellular and synaptic differentiation in the developing central nervous system. Thus, the deficit of $\alpha 5$, $\beta 3$ and $\gamma 3$ subunits is likely to cause deleterious effects such as the tremor and jerky gait that appear in p^{cp} mice. Moreover, it is possible that a $\beta 3$ deficit leads to other development anomalies associated with p^{cp} (cleft palate and runting for example), because $\beta 3$ is expressed in germinal zones¹⁶ and the mitotic zone of the forebrain²². Alternatively, the absence of other, as-yet unidentified, gene(s) in the p^{cp} deletion may be responsible for these phenotypes. Analysis of temporal and spatial expression patterns of GABA_A receptor subunit, together with histological analysis of the developing p^{cp} brain, could provide an insight into the neurological and cellular basis of p^{cp} phenotypes, and also clarify the significance of $\alpha 5$, $\beta 3$ and $\gamma 3$ subunits in development.

The human counterpart of the region deleted in p^{cp} is associated with Angelman syndrome (AS)⁵⁻⁹, which is characterized by severe mental retardation, microcephaly, seizures, ataxia, craniofacial anomalies and hypopigmentation^{23,24}. The smallest

known maternal deletion resulting in AS involves the $\beta 3$ but not the $\alpha 5$ gene^{8,9} (although a single AS patient bearing a translocation is apparently intact for the $\beta 3$ gene²⁵). Because the phenotypic effects of the p^{cp} mutation are recessive and independent of parental origin², the AS critical region may be outside of the p^{cp} deletion, despite certain phenotypic similarities between p^{cp} mice and AS patients¹⁵. However, if the mouse counterpart of the AS gene(s) is not imprinted, it is possible that the AS critical gene(s) may be within the p^{cp} deletion. Indeed, an AS-like paternal imprinting effect was not detected for the central region of mouse chromosome 7 (including the region deleted in the p^{cp} mutation²⁶). Conservation of synteny of the region deleted in p^{cp} with human chromosome 15q predicts that the human $\gamma 3$ gene will map near the $\alpha 5$ and $\beta 3$ genes. It thus remains to be determined what role, if any, these three GABA_A receptor subunit genes play in the aetiology of AS.

Note added in proof: Culiati *et al.*³² have reported a concordance of $\beta 3$ gene disruption and cleft palate phenotype in a series of p allele deletions. □

Received 12 March; accepted 21 May 1993.

1. Silvers, W. K. *The Coat Colors of Mice* 90–101 (Springer, New York, 1979).
2. Lyon, M. F. *et al. Proc. natn. Acad. Sci. U.S.A.* **89**, 6968–6972 (1992).
3. Phillips, R. J. *Mouse Newsletter* **56**, 18 (1977).
4. Khreshchatsky, M. *et al. Neuron* **3**, 745–753 (1989).
5. Wagstaff, J. *et al. Am. J. hum. Genet.* **49**, 330–337 (1991).
6. Hamabe, J. *et al. Am. J. med. Genet.* **41**, 54–63 (1991).
7. Saitoh, S. *et al. Lancet* **339**, 366–367 (1992).
8. Sinnott, D. *et al. Am. J. hum. Genet.* **52**, 1216–1229 (1993).
9. Knoll, J. H. M. *et al. Hum. molec. Genet.* **2**, 183–189 (1993).
10. DeLorey, T. & Olsen, R. W. *J. biol. Chem.* **267**, 16747–16750 (1992).
11. Nicholls, R. D. *Mouse Newsletter* **84**, 87–88 (1989).
12. Chaillet, J. R., Knoll, J. H. M., Horsthemke, B. & Lalonde, M. *Genomics* **11**, 773–776 (1991).
13. Wagstaff, J., Chaillet, J. R. & Lalonde, M. *Genomics* **11**, 1071–1078 (1991).
14. Nakatsu, Y., Gondo, Y. & Brilliant, M. H. *Mammal. Genome* **2**, 69–71 (1992).
15. Brilliant, M. H. *Mammal. Genome* **3**, 187–191 (1992).
16. Laurie, D. J., Wisden, W. & Seeburg, P. H. *J. Neurosci.* **8**, 4151–4172 (1992).
17. Wilcox, A. S. *et al. Proc. natn. Acad. Sci. U.S.A.* **89**, 5857–5861 (1992).
18. Gardner, J. M. *et al. Science* **257**, 1121–1124 (1992).
19. Poulter, M. O., Baker, J. L., O'Carroll, A. M., Lolait, S. J. & Mahan, L. C. *J. Neurosci.* **8**, 2888–2900 (1992).

20. Pritchett, D. B. & Seeburg, P. H. *J. Neurochem.* **54**, 1802–1804 (1990).
21. Korpi, E. R., Kleingor, C., Kettenmann, H. & Seeburg, P. H. *Nature* **361**, 356–359 (1993).
22. Zhang, J. H., Sato, M. & Tohyama, M. *J. comp. Neurol.* **308**, 586–613 (1991).
23. Angelman, H. *Dev. Med. Child. Neurol.* **7**, 681–683 (1965).
24. Magenis, R. E. *et al. Am. J. med. Genet.* **35**, 333–349 (1990).
25. Reis, A. *et al. Lancet* **341**, 122–123 (1993).
26. Cattanaach, B. M. *et al. Nature Genet.* **2**, 270–274 (1992).
27. Wilson-Shaw, D., Robinson, M., Gambarana, C., Siegel, R. E. & Sikela, J. M. *FEBS Lett.* **284**, 211–215 (1991).
28. Kirkness, E. F. *et al. Genomics* **10**, 985–995 (1991).
29. Ymer, S. *et al. EMBO J.* **8**, 1665–1670 (1989).
30. Rappolee, D. A., Mark, D., Banda, M. J. & Werb, Z. *Science* **241**, 708–712 (1988).
31. Olsen, R. W., McCabe, R. T. & Warmley, J. K. *J. chem. Neuroanat.* **3**, 59–76 (1990).
32. Culiati, C. T. *et al. Proc. natn. Acad. Sci. U.S.A.* **90**, 5105–5109 (1993).

ACKNOWLEDGEMENTS. We thank K. Tartof and C. Emerson for helpful comments on the manuscript. This work was supported by grants from the NIH (M.H.B. and should be A.J.T.); the Commonwealth of Pennsylvania and by the Pew Charitable Trust (M.H.B.). Y.N. was the recipient of a Uehara Memorial Fellowship. R.F.T. is the recipient of a fellowship from the Medical Research Council of Canada.

Amplification of calcium-induced gene transcription by nitric oxide in neuronal cells

Natalia Peunova & Grigori Enikolopov*

Cold Spring Harbor Laboratory, PO Box 100, Cold Spring Harbor, New York 11724, USA

NITRIC oxide (NO) is a short-lived, highly reactive gas, which has been identified as a mediator in vasodilation, an active agent in macrophage cytotoxicity and neurotoxicity, and a neurotransmitter in the central and peripheral nervous systems¹⁻⁵. Production of NO by neurons is critical for facilitated synaptic transmission in models of synaptic plasticity such as long-term potentiation and long-term depression, suggesting a role for NO as a retrograde messenger that could complete a hypothetical feedback loop by strengthening the connection between postsynaptic and presynaptic cells⁶⁻¹⁰. We report here that although alone NO has no evident effect on transcription, it can act as an amplifier of calcium signals in neuronal cells. NO and Ca²⁺ action have to coincide in time for amplification to occur. Experiments with a series of simplified reporter genes in combination with specific recombinant protein kinase inhibitors suggest that induction of

gene activity following NO-amplified calcium action involves protein kinase A-dependent activation of the transcription factor CREB.

To model a situation in which a neuron receives a NO signal simultaneously with a signal of a different modality, we exposed neuronal PC12 cells to combinations of NO with different inducers of transcription and monitored changes in *c-fos* and *c-jun* messenger RNAs. These immediate-early genes serve as sensitive reporters¹¹, the modular structure of their promoter regions allowing correlation of incoming signals with discrete transcriptional regulatory elements and factors. Further, these genes encode components of a family of transcription factors that may mediate subsequent long-term changes in patterns of cellular gene expression¹¹. Figure 1a shows that NO (produced by a nitric oxide-generating compound, sodium nitroprusside, SNP) was unable to induce transcription on its own; this holds true over a wide range (from 0.1 μ M to 1 mM) of SNP concentrations (data not shown). But when NO was applied to cells in combination with certain treatments known to induce immediate-early gene expression, induction of *c-fos* expression was greatly enhanced in some cases. NO only amplified the action of those agents that act through calcium ions — ionophores A23187 and ionomycin, Bay K8644 (an agonist of voltage-gated calcium channels (VGCC), thapsigargin (an agonist of intracellular calcium release) and KCl, which depolarizes the cell membranes. In contrast, NO did not affect induction by forskolin or phorbol myristate acetate (PMA). NO had similar effects on *c-jun* induction, although to a much lesser degree. NO-mediated amplification was blocked by the Ca²⁺ chelator EGTA, by the VGCC

* To whom correspondence should be addressed.

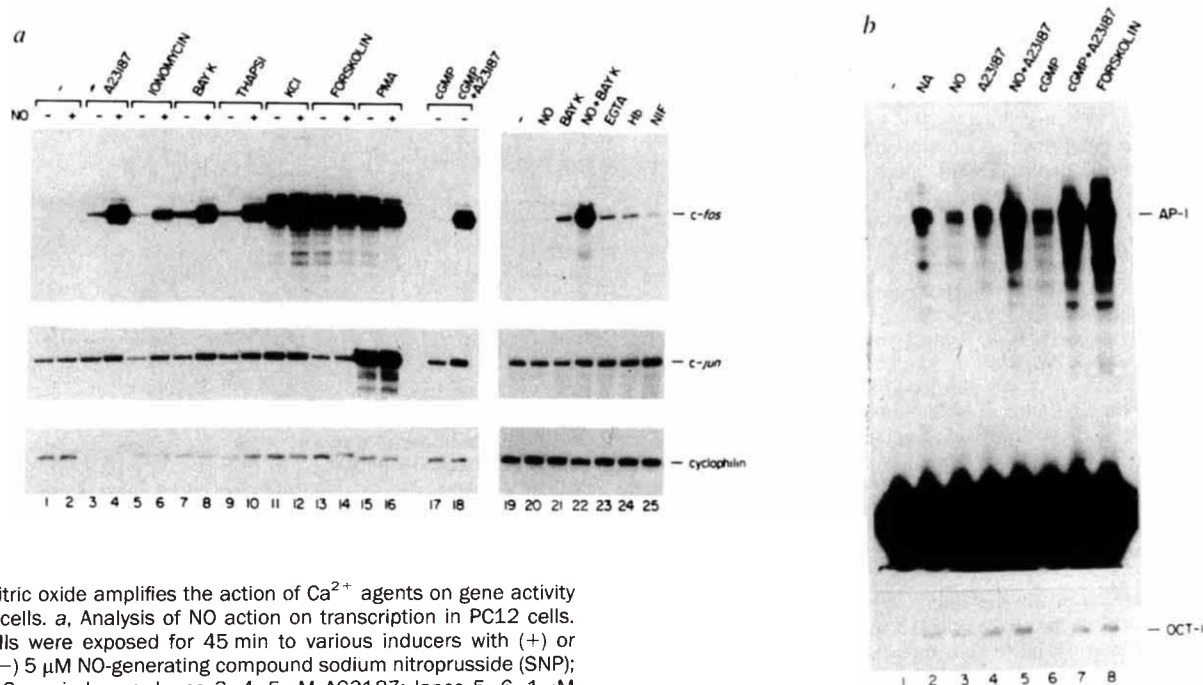


FIG. 1 Nitric oxide amplifies the action of Ca^{2+} agents on gene activity in PC12 cells. **a**, Analysis of NO action on transcription in PC12 cells. PC12 cells were exposed for 45 min to various inducers with (+) or without (–) 5 μM NO-generating compound sodium nitroprusside (SNP); lanes 1, 2, no inducers; lanes 3, 4, 5 μM A23187; lanes 5, 6, 1 μM ionomycin; lanes 7, 8, 5 μM Bay K8644; lanes 9, 10, 1 μM thapsigargin; lanes 11, 12, 55 mM KCl; lanes 13, 14 10 μM forskolin and 0.5 mM isobutylmethylxanthine (IBMX); lanes 15, 16, 200 ng ml^{-1} PMA; lane 17, 20 μM 8-Br-cGMP; lane 18, 20 μM 8-Br-cGMP and 5 μM A23187. Lanes 19–25 represent a separate set of experiments where the cells received no treatment (lane 19), 5 μM SNP (lane 20), 5 μM Bay K8644 (lane 21) or a combination of both (lanes 22–25), but were also treated with 3 mM EGTA (lane 23), 500 $\mu\text{g ml}^{-1}$ of haemoglobin (lane 24) and 20 μM nifedipin (lane 25) 5 min before addition of SNP and Bay K8644. **b**, Enhancement of AP-1 binding activity by combined NO/ Ca^{2+} action. Nuclear extracts were prepared from untreated cells (lane 2) and from cells treated with 5 μM SNP (lane 3), 5 μM A23187 (lane 4), a combination of both (lane 5), 20 μM 8-Br-cGMP without (lane 6) or with (lane 7) 5 μM A23187, or 10 μM forskolin and 0.5 mM IBMX (lane 8). Nuclear extracts were incubated with ^{32}P -labelled probes corresponding to the AP-1 binding site (AP-1) or the octamer motif (Oct-1) from the HSV ICPO promoter.

METHODS. PC12 cells were grown in DMEM supplemented with 5% calf serum and 10% horse serum (HyClone). For induction, cells were exposed for 45 min to various inducers in the presence or absence of 5 μM SNP, dissolved in H_2O immediately before addition. Cells were collected on ice, washed with ice-cold PBS and the cytoplasmic RNA was extracted and analysed by RNase protection following standard procedures. The *c-fos* probe was prepared from a plasmid carrying mouse *c-fos* sequences from –56 to +109. The *c-jun* probe contained

rat *c-jun* sequences from +295 to +527; cyclophilin probe (a gift from G. D'Arcangelo) contained rat cyclophilin sequences from +465 to +673. Rat *c-fos* mRNA carries three mismatches with the mouse probe, but under our T1 RNase digestion conditions it protected an RNA fragment of the expected length. The enhancement effect was reproduced by different sources of NO, including S-nitrosocysteine and 3-morpholino-sydnominine (data not shown), suggesting that this effect is specific to NO. For band-shift analysis, PC12 cells were stimulated for 90 min, washed with ice-cold PBS and the nuclei were isolated by treatment with 0.1% NP-40. Nuclear proteins were extracted in high-salt buffer and binding assays were as described¹⁴. Synthetic oligonucleotides contained a single copy of AP-1 site (5'-TCGACGGTATCGATA-GCTATGACTCATCCGGGGGATC-3') (a gift from K. Riabowol). Experiments with unlabelled wild-type and mutant AP-1 binding oligonucleotides and Fos- and Jun-specific antibodies confirmed that this elevation was due to specific binding by Fos- and Jun-containing protein complexes (data not shown). Treatment of PC12 cells with forskolin and IBMX produced the biggest increase in AP-1 binding activity and is reproduced (b, lane 8) for comparison with other inducers. For Oct-1 assays the same conditions were used except that each reaction contained 0.1 vol. of fetal bovine serum and the probe used was a 120 bp fragment containing a TAATGARAT element derived from the HSV ICPO promoter (a gift from J. S. Lai).

antagonist nifedipin, and by extracellular haemoglobin, which binds NO and sequesters it from the media (lanes 23–25). The NO/ Ca^{2+} induction of the endogenous *c-fos* gene and a transfected *fos* promoter showed rapid and transient kinetics characteristic of *c-fos* activation by a variety of inducers^{11,12} (Fig. 2a).

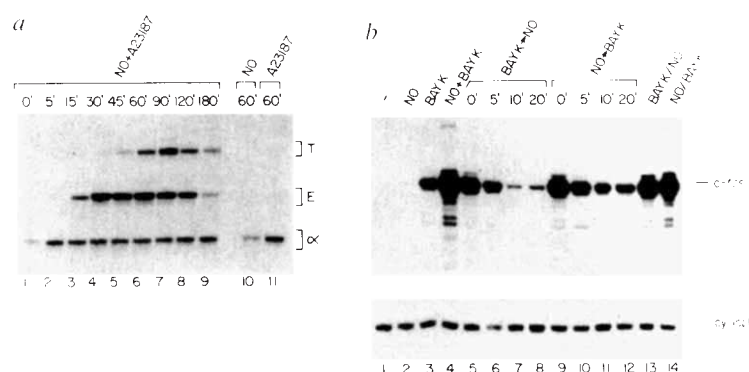
The elevated induction of *c-fos* mRNA was productive, leading to an increase in binding activity of the AP-1 transcription factor comprising Fos and Jun family proteins. The combined action of NO and the Ca^{2+} ionophore A23187 produced a strong increase in AP-1 activity (Fig. 1b).

To test the possibility that NO acted by changing the level of cellular cyclic GMP¹⁻⁵, we treated the cells with a combination of A23187 and 8-Br-cGMP. Application of 8-Br-cGMP did not induce transcription, but, like NO, 8-Br-cGMP amplified the effect of Ca^{2+} influx (Fig. 1a, lanes 17, 18). This effect was also reproduced when AP-1 binding was measured (Fig. 1b). These data support the notion that cGMP may mediate NO action and indicate that cGMP can influence gene expression. Taken

together, these results suggest that the combined action of NO and calcium leads to production of a highly active transcription factor capable of reprogramming the pattern of gene expression in neurons.

To determine whether NO and calcium act coordinately to produce a synergistic effect on transcription, we gave a pulse of one component, washed it away and added the second component immediately or after a further incubation of 5, 10 or 20 min. Addition of Bay K8644 after the removal of NO resulted in a loss of the synergistic effect (Fig. 2b, lanes 9–12). This was observed even when application of the two inducers was separated by only 5 min or less. After 10 min the signal was back to the levels of induction by Bay K8644 alone (compare lanes 3, 11 and 12). In the reciprocal experiment, in which Bay K8644 was applied first, a similar loss in the synergistic action of the two signals was observed (lanes 5–8). As a control, if the second reagent was added after 10 min without removing the first, the amplified response was nearly unaffected (lanes 13, 14). These

FIG. 2 Time requirements for the synergistic NO/Ca²⁺ action. *a*, Kinetics of induction of endogenous and transfected *fos* genes after NO/Ca²⁺ action. Reporter constructs containing CAT gene fused to the -356/+109 promoter region of *c-fos* gene (-356-*fos*-CAT) and a human α -globin gene used as an internal control were transfected into PC12 cells. After induction by 5 μ M SNP and 5 μ M A23187, RNA was isolated at various time points, indicated on the figure, and analysed by RNase protection assay as described for Fig. 1. T, E and α mark the positions of probe fragments protected by correctly initiated transcripts of the transfected *c-fos*-CAT plasmid, endogenous *c-fos* RNA and transcripts of the α -globin internal control, respectively. *b*, Analysis of the time window of the synergistic NO/Ca²⁺ action. PC12 cells were left untreated (lane 1), or were exposed to 5 μ M SNP, 5 μ M Bay K8644, or a mixture of SNP and Bay K8644 (lanes 2, 3 and 4, respectively). In a separate series of experiments (lanes 5–8) PC12 cells received a pulse of 5 μ M SNP for 10 min, were washed several times with warm equilibrated DMEM, covered with warm equilibrated DMEM and, after indicated periods of time, received 5 μ M Bay K8644 for 45 min; in reciprocal experiments (lanes 9–12) cells received a 10 min pulse of 5 μ M Bay K8644, followed by washing, incubation and addition of 5 μ M SNP after indicated periods of time. Lane 13, Bay K8644 (5 μ M) was added to cells for 10 min followed by addition of 5 μ M SNP for 45 min without removal of Bay K8644; lane 14, SNP (5 μ M) was added to cells followed by addition of 5 μ M Bay



K8644 without removal of SNP. RNA was isolated and assayed for *c-fos* RNA as described for Fig. 1.

METHODS. Plasmids were transfected into PC12 cells by electroporation at 220 V, 960 μ F. After a 40-h incubation the cells were stimulated for the indicated period of time, collected, and RNA was isolated as described in Fig. 1, except that RNA preparations were treated with RNase-free DNase and purified by deproteinization. RNA was analysed by RNase protection assay using *c-fos* and α -globin-specific probes as described for Fig. 1. The α -globin probe contained human α -globin gene sequences from -15 to +95.

observations indicate that the two stimuli, NO and calcium, have to be present together within a small time window to exert their synergistic action.

Calcium induction of a transfected reporter gene composed of the *c-fos* promoter linked to the heterologous chloramphenicol acetyltransferase coding sequences (*c-fos*-CAT)^{13,14} can also be amplified by NO, indicating that a short (400-base-pair (bp)) promoter region of the *c-fos* gene is sufficient to confer NO/Ca²⁺ inducibility (Fig. 2a). To examine which DNA sequences in the

c-fos promoter and which transcription factors mediate this effect, we tested a series of simplified reporter constructs^{13,14} (a gift from M. Gilman) for their ability to support NO-amplified calcium signalling. Promoter deletion analysis (Fig. 3b) shows two major transition points in induction levels, first when the reporter promoter was truncated from -356 to -275, which removed several important regulatory elements, and secondly when it was truncated from -71 to -56, removing the major cAMP-responsive element (CRE). These results are consistent

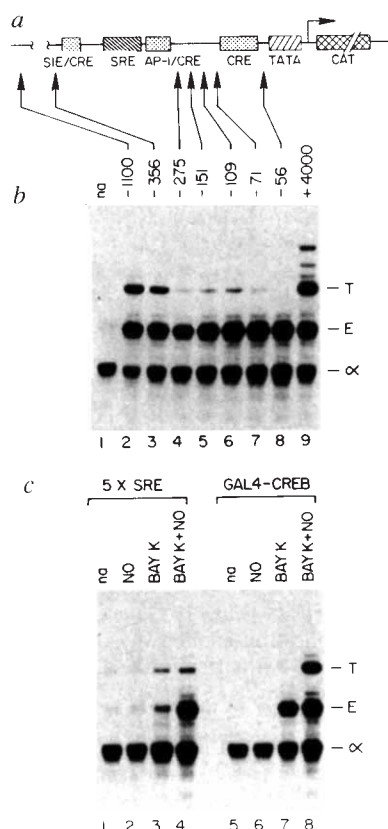


FIG. 3 CRE/CREB can transmit the amplified NO/Ca²⁺ signal to a reporter gene. *a*, Schematic diagram of *c-fos* upstream sequences. The regions corresponding to the *sis*-inducible element (SIE/CRE), the serum response element (SRE), the AP-1 binding site (AP-1/CRE), the cAMP response element (CRE), TATA box (TATA) and the CAT reporter gene are shown as filled boxes. *b*, Transient expression assay of mouse *c-fos* promoter carrying sequences up to +109 and connected to the coding region of CAT gene¹⁴ were transfected into PC12 cells along with the human α -globin internal control plasmid. Numbers above the lanes show the positions of 5'-end points of deletions of the *c-fos* promoter in *c-fos*-CAT fusions; lane 9 contained *c-fos* gene from position -356 to +4,000. *c*, Transient expression assay of multimerized SRE and of the GAL4/CREB reporter system. Cells were transfected with a reporter construct containing five copies of SRE connected to position -56 of *c-fos* promoter of the *c-fos*-CAT construct (5 $\times$ SRE, lanes 1–4), or with a mixture of 3 μ g of reporter plasmid 5 $\times$ GAL4-CAT and 0.3 μ g of expressor plasmid GAL4-CREB (GAL4-CREB, lanes 5–8) or its mutated (Ser 133 $\rightarrow$ Ala) version (lanes 9–12). The constructs were the same as in ref. 20, except that GAL4-CREB fusion was expressed under the control of RSV promoter. All transfections contained α -globin internal control plasmid. Cells were transfected and RNA was assayed as described in Fig. 2. Series of truncated *c-fos*-promoter, 5 $\times$ SRE-CAT, 5 $\times$ GAL4-CAT and GAL4-CREB constructs were gifts from M. Gilman and L. Berkowitz.

with the function of the *c-fos* CRE as one, but not the sole, Ca^{2+} response element (refs 15, 16 and N.P., M. Gilman and G.E., manuscript in preparation). To test whether the *c-fos* serum response element (SRE), which can also transmit part of a Ca^{2+} -generated signal in KCl-depolarized PC12 cells (N.P. *et al.*, manuscript in preparation), is a target for the NO/ Ca^{2+} pathway, we transfected cells with a reporter construct containing five copies of the SRE linked to a truncated *c-fos* promoter devoid of inducible sequences. Data in Fig. 3c (lanes 1–4) demonstrate that NO failed to amplify the calcium-dependent activation of this reporter. Thus, the SRE is not a target for NO-amplified Ca^{2+} signals.

CREs of several genes are bound by the transcription factor CREB^{17, 20}. To test directly the involvement of CREB as a downstream target of NO potentiation and to discriminate CREB-mediated effects from the action of endogenous transcriptional factors, we used effector proteins in which the DNA-binding specificity of CREB has been changed by fusion to the DNA-binding domain of the yeast transcriptional activator GAL4²⁰. A reporter gene containing multiple GAL4 binding sites served as a sensitive indicator of the transcriptional activity of the hybrid CREB protein. The combined action of NO and Bay K8644 synergistically activated the reporter gene (Fig. 3c). A mutant form of GAL4-CREB lacking the Ser 133 phosphoacceptor site could no longer transmit the NO signal, thus demonstrating that a functional CREB protein is a target for at least one of the NO/ Ca^{2+} -induced pathways in the nucleus.

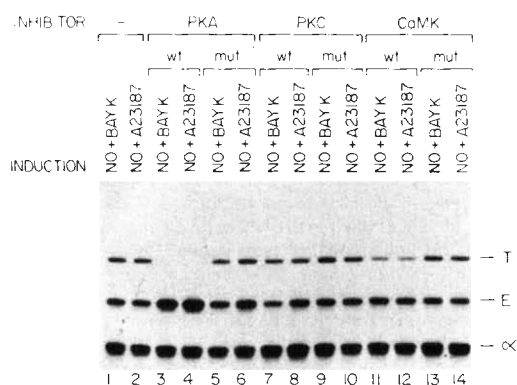


FIG. 4 Analysis of input of different protein kinases in NO-amplified calcium action. PC12 cells were cotransfected with -356-fos-CAT reporter construct along with 20 μg of recombinant inhibitors of different protein kinases and human α -globin control. Cells were exposed for 60 min to the action of 5 μM SNP and 5 μM A23187, or 5 μM SNP and 5 μM Bay K8644, as indicated, and RNA was isolated and assayed as described on Fig. 2. Reporter constructs were cotransfected either with pUC119 (lanes 1, 2) or with active (wt) and mutated inactive (mut) versions of recombinant protein kinase inhibitors as indicated. Lanes 3–6, recombinant inhibitors for protein kinase A (PKA); lanes 7–10, inhibitors for protein kinase C (PKC); lanes 11–14, inhibitors for Ca^{2+} /calmodulin-dependent multifunctional protein kinase II (CaMK). Recombinant inhibitors of protein kinases were expressed under the control of RSV promoter and rabbit β -globin splicing/polyadenylation sites. The structures of the recombinant PKC and CaMK inhibitors were matched to the sequences of the short conserved peptides located at positions 19–36 in various forms of PKC and positions 273–302 of CaMK, respectively. The recombinant PKA inhibitors included sequences of the murine PKA regulatory subunit $\text{RI}\alpha$ and the natural heat-stable PKA inhibitor (PKI). Such peptides act as powerful and highly specific pseudosubstrate inhibitors of the cognate enzymes *in vitro*^{24, 26}. Mutant forms of the inhibitors carrying inactivating amino-acid substitutions in the kinase recognition sequences served as controls. In a series of control experiments (N.P. *et al.*, manuscript in preparation and data not shown) the actions of recombinant inhibitors was shown to be selective for their target kinases and potent enough to block the reporter induction by a corresponding stimuli (forskolin, TPA, KCl).

Ca^{2+} signalling in neurons is mediated by several serine/threonine-specific protein kinases^{21, 23}. To determine which enzymes are involved in signalling by NO/ Ca^{2+} , we have used specific recombinant inhibitors of individual protein kinases (N.P. *et al.*, manuscript in preparation). These inhibitors were constructed based on the autoinhibitory pseudosubstrate domains of several protein kinases^{24, 26} and they are potent and selective toward their cognate enzymes (N.P. *et al.*, manuscript in preparation). We cotransfected reporter plasmids carrying a full-length *c-fos* promoter with inhibitor plasmids and monitored whether selective inhibition of a particular protein kinase blocked NO/ Ca^{2+} induction. Mutant inactive forms of the inhibitors served as controls. Consistent with the data on promoter mapping, only the cAMP-dependent protein kinase (PKA) inhibitor and, to a small degree, Ca^{2+} /calmodulin-dependent protein kinase II (CaMK) inhibitor blocked the NO/ Ca^{2+} response (Fig. 4). These results implicate the PKA-CREB-CRE system as a major component of the signalling pathway for the transcriptional synergy of NO and Ca^{2+} and suggest that cAMP- Ca^{2+} synergism^{15, 27} may be a part of NO/ Ca^{2+} signalling.

The phenomenon of NO-mediated potentiation of the Ca^{2+} response may have implications beyond transcriptional regulation. For example, NO could potentiate the PKA phosphorylation of cytoplasmic proteins with direct roles in synaptic function. The effect of NO on signalling might be particularly important at very low levels of calcium action, at which this inducer acting alone would have negligible effect; these very weak signals, which would go unnoticed by the cell, might be amplified by NO, resulting in pronounced physiological changes for the cell. Because NO diffuses freely, nearby synapses that receive very weak impulses simultaneously with exposure to NO might establish facilitated synaptic transmission²⁸. NO and Ca^{2+} have to act within a very narrow time window for this enhancement to occur, suggesting that in the nervous system, this synergistic effect might be restricted to the recently active synapses, thereby coinciding with transient elevations of calcium levels²⁹. □

Received 21 December 1992; accepted 28 May 1993.

- Moncada, S., Palmer, R. M. J. & Higgs, E. A. *Pharmac. Rev.* **43**, 109–142 (1991).
- Nathan, C. F. & Hibbs, J. B. *Curr. Opin. Immun.* **3**, 65–70 (1991).
- Garthwaite, J. *Trends Neurosci.* **14**, 60–67 (1991).
- Snyder, S. H. *Science* **257**, 494–496 (1992).
- Bredt, D. S. & Snyder, S. H. *Neuron* **8**, 3–11 (1992).
- O'Dell, T. J., Hawkins, R. D., Kandel, E. R. & Arancio, O. *Proc. natn. Acad. Sci. U.S.A.* **88**, 11285–11289 (1991).
- Schuman, E. M. & Madison, D. V. *Science* **254**, 1503–1506 (1991).
- Bohme, G. A., Bon, C., Stutzmann, J.-M., Doble, A. & Blanchard, J.-C. *Eur. J. Pharmac.* **199**, 379–381 (1991).
- Haley, J. E., Wilcox, G. L. & Chapman, P. F. *Neuron* **8**, 211–216 (1992).
- Shibuki, K. & Okada, D. *Nature* **349**, 326–328 (1991).
- Curran, T. in *Oncogene Handbook* (eds Reddy, E. P., Skalka, A. M. & Curran, T.) 307–325 (Elsevier, Amsterdam, 1988).
- Sheng, M., Dougan, S. T., McFadden, G. & Greenberg, M. E. *Molec. cell. Biol.* **8**, 2787–2796 (1988).
- Gilman, M. Z. *Genes Dev.* **2**, 394–402 (1988).
- Berkowitz, L. A., Riabowol, K. T. & Gilman, M. Z. *Molec. cell. Biol.* **9**, 4272–4281 (1989).
- Sheng, M., McFadden, G. & Greenberg, M. E. *Neuron* **4**, 571–582 (1990).
- Sheng, M. & Greenberg, M. E. *Neuron* **4**, 477–485 (1990).
- Gonzales, G. A. & Montminy, M. R. *Cell* **59**, 675–680 (1989).
- Gonzales, G. A. *et al.* *Nature* **337**, 749–752 (1989).
- Hoeffler, J. P., Meyer, T. E., Yun, Y., Jameson, J. L. & Habener, J. F. *Science* **242**, 1430–1433 (1988).
- Berkowitz, L. A. & Gilman, M. Z. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5258–5262 (1990).
- Malinow, R., Schulman, H. & Tsien, R. W. *Science* **245**, 862–865 (1989).
- Schulman, H. *Curr. Opin. Neurobiol.* **1**, 43–52 (1991).
- Edelman, A. M., Blumenthal, D. K. & Krebs, E. G. *Rev. Biochem.* **56**, 567–613 (1987).
- Soderling, T. R. *J. biol. Chem.* **265**, 1823–1826 (1990).
- Kemp, B. E., Pearson, R. B. & House, C. M. *Meth. Enzym.* **201**, 287–304 (1991).
- Walsh, D. A. & Glass, D. B. *Meth. Enzym.* **201**, 304–316 (1991).
- Nguyen, T. V., Kobierski, L., Comb, M. & Hyman, S. E. *J. Neurosci.* **10**, 2825–2833 (1990).
- Bonhoeffer, T., Kossel, A., Boiz, J. & Aertsen, A. *Cold Spring Harb. Symp. quant. Biol.* **55**, 137–146 (1990).
- Gally, J. A., Montague, P. R., Reeke, G. N. & Edelman, G. M. *Proc. natn. Acad. Sci. U.S.A.* **87**, 3547–3551 (1990).

ACKNOWLEDGEMENTS. We thank M. Gilman and J. Watson for support and encouragement; M. Gilman and L. Berkowitz for the gift of reporter constructs; K. Riabowol, J.-S. Lai and G. D'Arcangelo for reagents; and R. Davis, M. Gilman, W. Herr, A. Silva and B. Stillman for critically reading the manuscript. We greatly acknowledge the financial support from I. Hechler and from the Donaldson Charitable Trust arranged by W. E. Murray.

Dimerization and the control of transcription by *Krüppel*

Frank Sauer & Herbert Jäckle

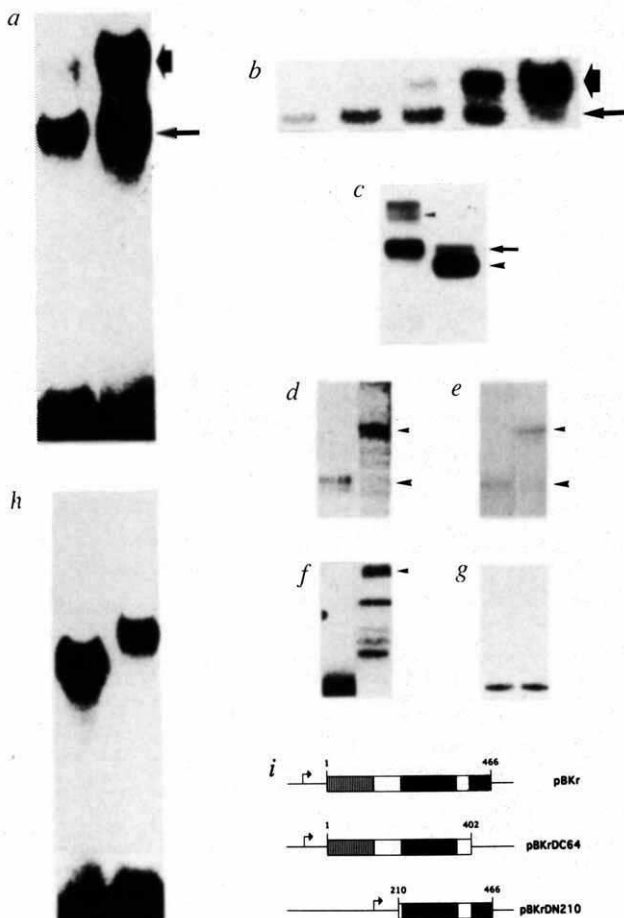
Abt. Molekulare Entwicklungsbiologie, Max-Planck-Institut für biophysikalische Chemie, Postfach 2841, D-37018 Göttingen, Germany

KRÜPPEL (KR), a *Drosophila* zinc finger-type¹ transcription factor^{2,4}, can both activate and repress gene expression through interaction with a single DNA-binding site⁴. The opposite regulatory effects of KR are concentration-dependent, and they require distinct portions of KR such as the N-terminal region for activation and the C-terminal region for repression⁴. Here we show that KR is able to form homodimers through sequences located within the C terminus. When these sequences were fused to separated functional parts of the yeast transcription factor GAL4⁵, they reconstituted a functional transcriptional activator on dimerization *in vivo*. Our results suggest that the KR monomer is a transcriptional activator. At higher concentration KR forms a homodimer and becomes a repressor that functions through the same target sequences as the activator.

A single KR *in vitro* binding site of the sequence AAAAGGGTTAA ('K-element')⁶ mediates activation or repression of chloramphenicol acetyl transferase (CAT) reporter gene expression in response to distinct concentrations of KR in cultured *Drosophila* Schneider cells⁴. Low levels of KR activate CAT gene expression, whereas high levels of KR lead to repression⁴. This suggests that different, concentration-dependent forms of the transcription factor KR result in opposite regulatory effects through interactions with a single KR binding

FIG. 1 Concentration-dependent homodimer formation of KR is mediated by the C-terminal sequences. **a**, Gel mobility experiment showing one and two DNA-protein complexes formed between *in vitro* translated KR and K-element DNA. Note a single KR-DNA complex (thin arrow) at low concentration of KR and the second, more slowly migrating complex (fat arrow) at 20-fold higher concentration. Free K-element DNA is seen at the bottom of the lane. Note that the K-element represents a single, non-repeated minimal KR-binding site, and that the DNase I-protected sequences were not altered in footprinting experiments under conditions when the KR monomer and homodimers were formed. Thus, we do not know whether the KR homodimer is bound to a single KR binding site, or whether the two DNA-binding domains each bind to a separate K-element. **b**, Gel mobility experiment showing that increasing amounts of *in vitro* translated KR (1, 5, 25, 125 and 625 µg KR; left to right) lead to a concentration-dependent increase of the more slowly migrating complex (fat arrow). **c**, A mixture of KR and DN210 (through cotranslation *in vitro*) leads to a third protein-DNA complex (small arrowhead) of intermediate size. In contrast, an intermediate complex between KR and DC64 could not be observed. A mixture containing low amounts of KR (arrow) and excess DC64 (arrow head) is shown on the right lane; note that equal amounts of KR and DC64, and mixtures containing high amounts of KR and decreasing amounts of DC64 (and vice versa) did not result in KR/DC64 complexes (data not shown). Thus, KR and DN210, but not KR and DC64, can form heterodimers as confirmed by crosslinking experiments (see **g**). **d**, Autoradiograph of SDS-PAGE of *in vitro* translated ³⁵S-methionine-labelled KR which was immunoprecipitated with anti-KR antibodies before (left lane) and after (right lane) cross-linking. Note KR (large arrowhead), and a band of twice the KR size (small arrowhead) when cross-linker was added. **e**, ³⁵S-Methionine-labelled KR selected through binding to biotin-labelled K-element DNA. Untreated (left lane) or cross-linker treated (right lane) *in vitro* translated KR was incubated with biotin-labelled K-element DNA, precipitated by streptavidin agarose, and analysed by SDS-PAGE. In the absence of cross-linking agent (left lane), KR (large arrowhead) was detected. In cross-linker-treated mixtures (right lane), a second protein band (small arrowhead) of twice the size of KR was observed. **f, g** Autoradiographs of SDS-PAGE of *in vitro* translated ³⁵S-methionine-labelled DN210 (**f**) or DC64 (**g**) which were immunoprecipitated with anti-KR-antibodies before (left) or after (right) cross-linking. Note dimerized DN210 (small arrowhead) in (**f**). In the absence of the C-terminal 64 amino acids (DC64) only a single band can be obtained with or without cross-linking (**g**). Note that dimer formation (**d, f**) requires the presence of the K-element. **h**, Single complexes formed between a truncated Kr protein lacking the N-terminal 210 amino acids (left lane) or the C-terminal 64 amino acids (right lane). Note that both at low and high protein concentrations only single complexes are formed. They contain a DN210 homodimer (**f**), and a DC64 monomer bound to the K-element DNA (**g**), indicating that homodimer formation requires the C-terminal 64 amino acids of KR. Note that KRDN64 monomer (402 amino acids) slows down the DNA migration stronger than the KRDN210 dimer (512 amino acids total). This effect might be caused by the difference of the overall charge and differences in the structure of the two truncated proteins (KRDN210 lacks one third of the charged amino acids and a series of prolines which are located within the N-terminal region¹). **i**, Schematic representation of fusion gene constructs used for *in vitro* transcription/translation experiments to synthesize KR and truncated versions of Kr protein. Plasmids pBKr, pBKrDC64 and pBKrDN210 contain the entire KR open reading frame (amino acids 1-466)¹ or the indicated portions fused to the T7-promoter (arrow) of Bluescript plasmid DNA (Stratagene). The KR DNA-binding domain (amino acids 221-333)¹ is shown by the black bar. pBKrDC64 gives rise to a truncated KR (DC64) which lacks amino acids 403-466¹ (grey bar). Note that the deleted 64 amino acids correspond to the region required for KR repressor function⁴. pBKrDN210 gives rise to a protein lacking the N-terminal 210 amino acids of KR (DN210) required for KR-dependent activation (striped bar)⁴.

METHODS. *In vitro* transcription and translation were according to the manufacturer's (Promega) instructions. For bandshift assays, aliquots of lysates programmed with Kr messenger RNA were incubated (25 min, room temperature) in a total reaction volume of 15 µl containing binding buffer B (25 mM HEPES, pH 7.9, 50 mM KCl, 7% glycerol, 0.1 mM ZnSO₄, 1 mg ml⁻¹ BSA), 1-5 fmol ³²P-end-labelled KR target DNA (a 21-bp oligonucleotide containing the 11-bp KR *in vitro* binding site)⁶, and 1 µg salmon sperm competitor DNA (Sigma), and samples were separated on a 6% native polyacrylamide gel¹⁶. Glutaraldehyde cross-linking was as described⁷, and the reaction mixtures containing ³⁵S-methionine-labelled protein were separated by SDS-PAGE⁷. For co-immunoprecipitation, cross-linker-treated or non-treated ³⁵S-methionine-labelled proteins were incubated with anti-KR-antibody¹⁷ in buffer B. Antibody-protein complexes were precipitated with protein A-sepharose and analysed by SDS-PAGE. For the precipitation of DNA-bound protein complexes, 4-6 µl of *in vitro* translated ³⁵S-methionine-labelled proteins were incubated in 20 µl binding buffer B with 20-100 fmol biotinylated KR target DNA (30 min on ice). DNA-protein complexes were precipitated by streptavidin agarose (Sigma) and analysed as described¹⁸.



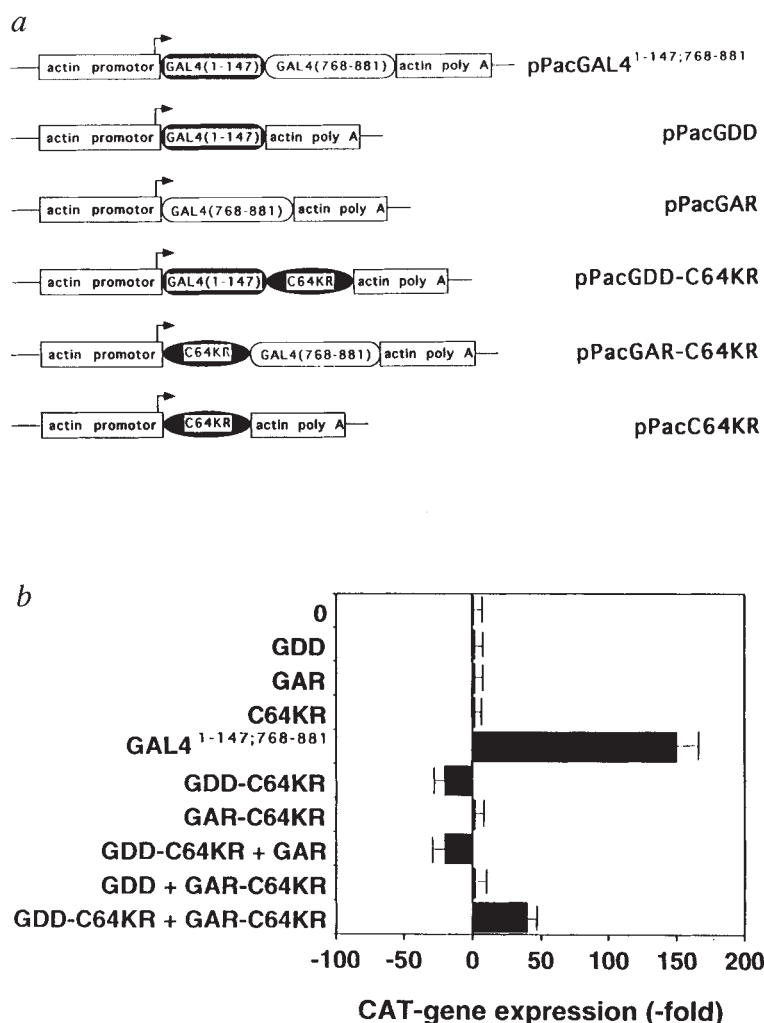
site. We therefore examined DNA-protein complex formation between KR and K-element DNA at different KR concentrations. Gel mobility shift assays with nuclear extracts from *Kr*-transfected cells, or *in vitro* translated KR, revealed two specific, KR-dependent DNA-protein complexes (Fig. 1a). The more rapidly migrating complex was found at low and high concentrations of KR, whereas the second more slowly migrating complex was critically dependent on the KR concentration relative to its target DNA (Fig. 1b). To determine whether the two complexes contain one and two KR molecules, respectively, we performed cross-linking⁷ combined with immunoprecipitation experiments^{8,9}, and gel mobility assays with KR and truncated *Kr* proteins. The experiments summarized in Fig. 1c-h indicate

that KR can form homodimers, and that the two DNA-protein complexes represent one or two KR molecules bound to the target sequence.

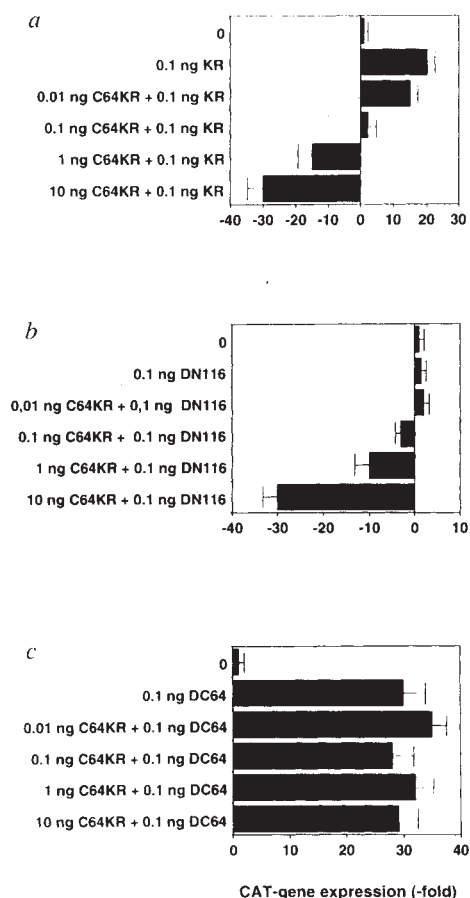
To delimit the region required for KR homodimer formation, we analysed N- and C-terminal truncated *Kr* protein⁴ termed DN210 and DC64 (Fig. 1i), respectively. DN210 can form both homodimers (Fig. 1f, h) and heterodimers with KR wild type (Fig. 1c), whereas DC64 is unable to form a dimer molecule (Fig. 1g, h). These results indicate that the protein sequences required for homodimer formation reside within the 64 amino acids of the KR C-terminus (C64KR) which are essential for KR repressor function⁴. To show C64KR-dependent protein dimerization by a functional assay, we used the GAL4 transcrip-

FIG. 2 The C64KR domain joins functional GAL4 domains by dimerization *in vivo*. **a**, Yeast transcription factor GAL4-derived gene constructs used for cotransfection experiments with the CAT reporter gene plasmid pADH33-1UAS that contains a single GAL4 DNA target sequence in front of the *Drosophila Adh* distal promoter^{19,20}. The effector plasmids contain the constitutive actin 5C promoter fused to parts of the GAL4 coding sequence (numbers refer to GAL4 amino-acid sequence⁵ maintained in the construct) to which the 3'-untranslated region of the actin 5C gene including the polyadenylation signal had been fused²¹. pPacGAL1-147;768-881 encodes a functional transcription activator, that is, the GAL4 DNA-binding domain (position 1-147)⁵ is linked to a GAL4 activating region (position 761-881)⁵. pPacGDD and pPacGAR encode for the GAL4 DNA-binding domain (amino acids 1-147) and for the GAL4 activating region (amino acids 761-881), respectively. pPacGDD-C64KR and pPacGAR-C64KR represent fusion genes in which the KR C-terminal 64 amino acids (position 403-466)⁴ are fused in frame with GDD and GAR coding sequences, respectively. pPacC64KR encodes the C-terminal 64 amino acids of KR¹. **b**, Effector plasmid-dependent CAT gene expression showing that the KR C-terminal domain acts as a repressor and dimerization domain. *Drosophila* Schneider cells^{4,22} were cotransfected with the different effector plasmid DNA shown in **a** and a constant amount of the CAT reporter plasmid pADH33-1UAS. Note that expression of GDD, GAR, C64KR, GAR-C64KR or the combination of GDD and GAR-C64KR had no significant effect on the basal level of CAT gene expression. But expression of GDD-C64KR led to significant CAT gene repression both in the absence or presence of GAR expression. Expression of GAL1-147;768-881 led to CAT gene activation. Also, coexpression of GDD-C64KR and GAR-C64KR led to significant activation of CAT gene expression (although GDD-C64KR expression led to repression and GAR-C64KR had no significant effect on the basal level of expression). These results suggest that the C64KR domain functions as a repressor domain, and that sequences of this domain are able to join the functional parts of the yeast transcription factor, thereby reconstituting a functional transcriptional activator *in vivo*. Bars represent mean values of at least four independent experiments; standard deviation is indicated.

METHODS. Schneider tissue culture cells were cotransfected with 3 μ g of reporter plasmid and 2 μ g of effector plasmid DNA. Cell collection and the determination of reporter gene expression were done as described⁴. pPacGDD and pPacGAR were generated by cloning the blunt-ended *Hind*III-*Bam*HI fragment of pMA424²³ encoding the GAL4 DNA binding domain, and the end-filled *Eco*RI-*Hind*III fragment of pMA236⁵ encoding an acidic activator domain, respectively, into the end-filled *Bam*HI site of pPac²¹. pPacGAL1-147;761-881 was generated by insertion of the end-filled 1 kb *Hind*III fragment of pMA236 into the end-filled *Bam*HI site of pPac. pPacC64KR was constructed by fusing a 0.8 kb *Hind*III fragment from pck2b¹, encoding the C-terminal 64 amino acids of KR with an artificial start codon-linker into the *Bam*HI sites of Bluescript and pPac to give pBlueC64 and pPacC64KR, respectively. For generating pPacGDD-C64KR, a *Hind*III fragment from pck2b¹, encoding the C-terminal 64 amino acids of KR was inserted into the end-filled *Bam*HI site of pBlueGDD. The 1.2 kb *Hind*III fragment of this vector, after a filling reaction, was inserted into the end-filled *Bam*HI site of pPac to generate



pPacGDD-C64KR. For constructing pPacGAR-C64KR, the 580-bp *Eco*RI-*Hind*III of pMA236 was cloned into the appropriate sites of Bluescript (Stratagene) to generate pBlueGAR. A 0.8 kb blunt-ended, *Bam*HI-*Dde*I fragment from pBlueC64KR was inserted into the end-filled *Eco*RI site of pBlueGAR to generate pBlueGAR-C64KR. The 1.3 kb *Pst*I-*Hind*III fragment from pBlueGAR-C64 was end-filled and cloned into the blunt-ended *Bam*HI site of pPac to generate pPacGAR-C64KR. The reporter plasmid pADH33-1UAS was generated by inserting a 40 bp oligonucleotide from MH100²⁴, containing a single GAL4 binding site, through *Sal*I linker into the *Sal*I site of pADH33^{19,20}.



tion factor of the yeast *Saccharomyces cerevisiae*¹⁰. GAL4 contains two separable, functionally essential domains. The N-terminal DNA-binding domain of GAL4 (GDD; amino acids 1–147)⁵ binds to a specific upstream activator site (UAS_G), and in conjunction with a C-terminal acidic region (GAR; amino acids 768–881)⁵, the resulting GAL41–147; 768–881 fusion protein is able to activate UAS_G-mediated transcription⁵ (Fig. 2). We fused each of the separate domains to the C64KR sequences. DNA constructs expressing the resulting hybrid genes GDD-C64KR or GAR-C64KR (Fig. 2a), or the separate GAL4 domains GDD or GAR, or combinations of these constructs were introduced into Schneider cells, and assayed for their ability to regulate UAS-dependent CAT gene expression. GDD, GAR and GAR-C64KR had no regulatory effect. The GDD-C64KR hybrid caused significant repression (Fig. 2b). This suggests that C64KR, which is essential for KR repressor function⁴, also provides repressor function to the GAL4 DNA-binding domain, a phenomenon which is beyond the scope of our present study. But when both GDD-C64KR and the hybrid gene GAR-C64KR, but not GAR, are coexpressed, activation of CAT gene expression had been observed (Fig. 2). This suggests that the C64KR domain must be able to link the DNA-binding and activation domains of GAL4, thereby reconstituting a functional transcriptional activator upon dimerization.

We then tested whether coexpression of C64KR, with activating concentrations of KR, is sufficient to convert the activator KR into a repressor. For this, we transfected Schneider cells with a K-element-dependent CAT reporter gene⁴, and with gene constructs expressing C64KR, or C64KR together with KR or truncated KR proteins (Fig. 3). Neither C64KR expression itself, nor its combination with DC64 caused transcriptional repression. But strong repression of gene expression was found when C64KR was coexpressed in combinations with KR or KR from which the N-terminal region had been removed⁴ (Fig.

FIG. 3 Coexpression of the C64KR domain alters the regulatory function of KR through interaction with the KR C-terminal sequences. *Drosophila* Schneider cells were cotransfected with a constant amount (4 µg) of the CAT reporter plasmid pADH86–1K which contains a single K-element in front of the *Drosophila* ADH promoter and constant amounts of the effector plasmids pPacKr, pPacKrDN116 and pPacKrDC64⁴ as indicated, and increasing amounts of the effector plasmid C64KR. **a**, CAT gene expression in response to KR and increasing amounts of C64KR. Note that KR-dependent activation is altered into C64KR-dependent repression. **b**, DN116, a truncated version of KR, which lacks the KR activator region⁴, mediates C64KR-dependent repression of CAT gene expression. **c**, CAT gene expression in response to C64KR which lacks the C-terminal 64 amino acids (position 402–466)⁴ is left unaltered by increasing amounts of C64KR. Note that increasing amounts of C64KR by itself have no effect on the basal level of CAT gene expression and it is unable to interfere with C64KR-mediated CAT gene expression. Bars represent mean values of at least four independent experiments; standard deviation is indicated.

METHODS. Schneider cells were cotransfected with the indicated amounts of effector plasmids⁴ and a constant amount of the reporter plasmid pADH86–1K⁴ as described in Fig. 2. Reporter gene expression was analysed as described⁴.

3a, b). Thus, the C64 sequences of KR contain target sequences for C64KR to act as a repressor of KR-dependent activation. This suggests that dimerization and repression are intimately linked functions of the C64KR domain, and that the switch in KR regulatory function is likely to be caused by homodimer formation.

In the embryo, genetic analysis predicts that *Krüppel* functions as a transient activator of the target gene *knirps* in a region of low KR concentration⁶, but other known interactions within the *Drosophila* segmentation gene cascade¹¹ involve repression by *Krüppel* (reviewed in ref. 11). One known mechanism by which KR exerts its function is through competition with activators for the binding to overlapping binding sites^{12,13}. A second, unrelated mechanism emerged from *Drosophila* tissue culture where KR acts as a repressor of *hunchback*-dependent gene activation in the absence of detectable KR binding sites, an effect that requires an intact KR C-terminal region³. KR also exerts repressor function, as has been described in mammalian tissue culture², through a KR repressor domain localized within the alanine-rich N-terminal region² similar to other transcriptional repressors¹⁴. This domain fully overlaps the N-terminal region shown to be required for gene activation in the *Drosophila* tissue culture system⁴, suggesting the possibility that transcriptional regulation by KR may involve different cell-specific protein-protein interactions. In fact, such interactions might also allow for the opposite regulatory responses to different KR concentrations from a single binding site, a finding that has no precedent in studies with other transcription factors. We propose that the activation of gene expression by KR requires distinct interactions of the KR monomer with components of the primary transcription complex¹⁵. When the KR concentration increases, homodimer formation takes place and alters these interactions in a manner that results in repression instead of transcriptional activation. □

Received 18 January; accepted 18 May 1993.

- Rosenberg, U. et al. *Nature* **310**, 336–339 (1986).
- Licht, J. D., Grossel, M. J., Figge, I. & Hansen, U. *Nature* **346**, 76–79 (1990).
- Zuo, P. et al. *Genes Dev.* **5**, 254–264 (1991).
- Sauer, F. & Jäckle, H. *Nature* **353**, 563–566 (1991).
- Ma, J. & Ptashne, M. *Cell* **48**, 847–853 (1987).
- Pankratz, M. J., Hoch, M., Seifert, E. & Jäckle, H. *Nature* **341**, 337–340 (1989).
- Courey, A. J. et al. *Cell* **59**, 827–836 (1989).
- Benson, M. & Pirota, V. *EMBO J.* **6**, 1387–1392 (1987).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. in *Molecular Cloning* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1990).

10. Fields, S. & Song, O. *Nature* **340**, 245–246 (1989).
11. Pankratz, M. J. & Jäckle, H. in *The Development of Drosophila melanogaster* (eds Bate, M. & Martinez-Arias, A.) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1993).
12. Small, S. et al. *Genes Dev.* **5**, 827–839 (1991).
13. Stanojevic, D., Small, S. & Levine, M. *Science* **254**, 1385–1387 (1991).
14. Han, K. & Manley, L. *Genes Dev.* **7**, 491–503 (1993).
15. Roeder, R. G. *Trends biochem. Sci.* **16**, 402–408 (1991).
16. Sorger, P. K. & Pelham, H. R. B. *EMBO J.* **6**, 3035–3041 (1987).
17. Gaul, U. & Jäckle, H. *Cell* **50**, 639–647 (1987).
18. Ingraham, H. A. et al. *Cell* **61**, 1021–1033 (1990).
19. Heberlein, U., England, B. & Tjian, R. *Cell* **41**, 965–977 (1985).
20. Heberlein, U. & Tjian, R. *Nature* **331**, 410–415 (1988).
21. Krasnow, M. A., Saffman, E. E., Kornfeld, K. & Hogness, D. S. *Cell* **57**, 1031–1043 (1989).
22. Schneider, I. J. *Embryol. exp. Morphol.* **27**, 353–365 (1972).
23. Ma, J. & Ptashne, M. *Cell* **51**, 113–119 (1987).
24. Kakidani, H. & Ptashne, M. *Cell* **52**, 161–167 (1988).

ACKNOWLEDGEMENTS. We thank E. Kaufmann, R. Rivera-Pomar and G. Chalepakos for critical discussions, I. Gajewski for photography, and D. Ruden for making us familiar with the GAL system. The work was supported by the Max-Planck-Gesellschaft and by the Fonds der Chemischen Industrie.

Force–velocity relationships in kinesin-driven motility

Kirsten Hall, Douglas G. Cole, Yin Yeh,
Jonathan M. Scholey* & Ronald J. Baskin

Biophysics Graduate Group, Section of Molecular and Cellular Biology, and Department of Applied Science, University of California, Davis, California 95616, USA

KINESIN is a microtubule-based motor protein that uses energy released from Mg-ATP hydrolysis to generate force for the movement of intracellular membranes towards the fast-growing (plus) ends of microtubule tracks in cells¹. Kinesin-driven microtubule movement can be visualized and quantified using light microscope motility assays^{2–5} but our understanding of how kinesin generates force and motion is incomplete⁶. Here we report the use of a centrifuge microscope^{7,8} to obtain force–velocity curves for kinesin-driven motility and to estimate that the maximal isometric force generated per kinesin is 0.12 ± 0.03 pN per molecule.

Purified kinesin uses Mg-ATP, in preference to other nucleotides, as a substrate for driving microtubule movement in a motility assay, with a Michaelis constant (K_m) of $60 \mu\text{M}$ ATP and an unloaded maximum velocity (V_{max}) of $0.6 \mu\text{m s}^{-1}$ (ref. 2). Individual kinesin molecules can drive this motility^{3,4} which proceeds along single protofilament tracks^{5,9}. The elements responsible for generating force and motion are located in the kinesin motor domain^{10,11}, but the basic features of mechanochemical coupling have not been described.

Using a centrifuge microscope, we monitored demembrated sea urchin sperm¹¹ being moved over the coverslip of a motility assay cell by highly purified sea urchin egg kinesin (Figs 1 and 2). The assay cell was inserted into the metal plate of the microscope, where it was rotated at speeds up to 5,000 r.p.m., and consequently the kinesin molecules moving sperm in the centrifugal and centripetal directions, respectively, experienced an increasing negative or positive force as the rate of rotation increased (see legend to Fig. 2). Increasing the centrifugal force consistently caused a decrease in the velocity of motility (Fig. 1a–d), but the observed force–velocity (f – v) curves were rectilinear (Fig. 1a, b) or curvilinear (Fig. 1c, d) depending upon the conditions of assay. We observed linear f – v curves when the measured ratio of ‘functional’ to total applied kinesin (the ‘specific functional density’) was relatively high, and curvilinear f – v curves when this ratio was relatively low (Fig. 1 legend).

In the presence of $\geq 10 \text{ mM}$ Mg-ATP, the unloaded velocity of movement was about $0.7 \mu\text{m s}^{-1}$; this value did not change much when the kinesin was diluted to a concentration below which we observed no motility. Figure 1a shows that under con-

ditions of higher specific functional density, the application of an increasing positive load (opposing the action of kinesin) to a level of ≥ 0.25 pN total load, resulted in a steady linear decrease in the velocity of microtubule movement until, at a load equal to the maximum isometric force, P_0 , movement stopped. Applying a negative load to the same preparation of kinesin did not increase the rate of motion as might be expected, but instead decreased in velocity of gliding (Fig. 1b), as reported previously with myosin⁸. These results suggest that ‘pushing’ a kinesin (or myosin) molecule in the direction that it is moving does not represent a simple reversal of increasing the load on it.

At a relatively low specific functional density of kinesin, curvilinear f – v curves (Fig. 1c, d), similar to those obtained previously with purified myosin⁸ and muscle fibres under positive loads¹², were obtained. In this case, some of the applied kinesin molecules appear to be inactivated during adsorption to the coverslip and may exert drag forces by binding non-productively to the moving microtubules leading to the deviations from linearity.

We have found that the maximal isometric force (P_0) increases in a saturable fashion with increasing densities of kinesin in the assay (Fig. 3). The results obtained below saturation support the hypothesis that the maximal isometric force is a function of the number of active motor molecules, as expected. The curve saturates at a point where the interaction of additional motor molecules with the sperm apparently does not increase the net force on it. From measurements of P_0 at the minimal kinesin density that would support movement, we estimate P_0 per molecule to be 0.12 ± 0.03 pN.

Our estimate of P_0 per kinesin is about 10–20-fold lower than the value obtained using ‘optical tweezers’¹³, but the reason for this discrepancy is unknown. We were concerned that a torquing force might ‘rip’ the axonemal microtubules off the kinesin-coated coverslip in the centrifuge microscope, artefactually lowering the measured force, but our analysis (Fig. 2 legend) suggests that this is unlikely; for example, the application of a positive centrifugal load would actually decrease the torque on the moving sperm. Of course, the existence of other unidentified sources of error in the measurement of force using the centrifuge microscope (or the optical tweezers for that matter) cannot be ruled out.

To understand fully the molecular basis of the f – v curves described in this report, further work will be needed. Our working hypothesis, however, is that the ‘working stroke’ of the kinesin motor domain may serve to ‘stretch’ a series elastic element whose ‘relaxation’ exerts force for moving microtubules. When this relaxation is opposed by increasing positive centrifugal loads, we can show that because the inertia is low, the velocity would fall in a linear manner as observed in Fig. 1a (to be described in detail elsewhere). Curvilinearity is thought to arise as a result of drag induced by motors that are bound to moving microtubules, but not exerting force on them. In our study, such motors are thought to be ‘inactive’ kinesin molecules that bind nonproductively to microtubules (for example, see Fig. 1c). In

* To whom correspondence should be addressed.

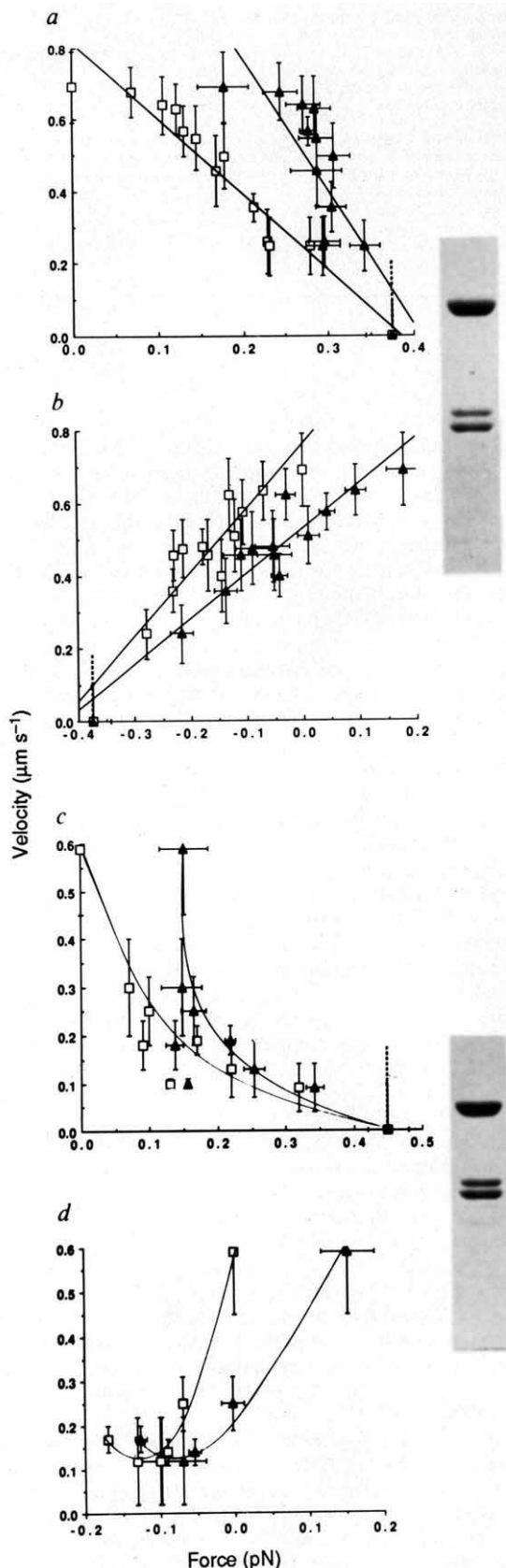


FIG. 1 Force-velocity curves for kinesin-driven motility. The rate of motility falls in a rectilinear fashion under some conditions (a, b) or in a curvilinear fashion under others (c, d) as the load is increased. Applied centrifugal force (□) and force corrected for viscous drag (▲). The data were obtained using the kinesin preparations shown in the Coomassie-

FIG. 2 Theory and calculations: Schematic representation of forces and torque acting on the motor molecule attached to the demembrated sperm which is shown with a conical head and a bent axoneme. The axoneme is assumed to be rigid so that the force at P is equal to the force at the junction of the axoneme and the head, which by geometric considerations, is close to the centre of mass of the sperm head (O). Because the mass of the axoneme is negligible in comparison with the mass of the sperm head, the centre of mass of the sperm will be approximately that of the sperm head alone. Both the motor force applied at P, F_m , and the centrifugal force, F_c , are assumed to be applied to the centre of mass of the sperm along the kinesin-coated surface in the (-x) and (+x) directions, respectively. The applied force, $F_c = \Delta\rho V\omega^2$, where the density of the sperm head (ρ) is assumed as usual to be 1.3 g ml^{-1} (although 1.6 g ml^{-1} has been reported¹⁶). Therefore, $\Delta\rho = 0.3$, sperm head volume $V = 8 \mu\text{m}^3$, rotor radius $r = 10.5 \text{ cm}$, and ω is the angular frequency of rotation of the stage. The net force along the x-axis $F_{\text{net}} = F_c - F_m$. The velocity of the sperm head is obtained from the equation of motion of a system with negligible inertial forces (Reynold's number is 2.1×10^{-7}). We write: $F_{\text{net}} = f v_x = F_c - F_m$, where f is the total frictional coefficient and v_x is the measured velocity along the +x direction. Thus, $v_x = (F_c - F_m)/f$. The unloaded velocity is obtained where $F_c = 0$. When $v_x = 0$, $F_c = F_m = P_0$, the isometric force. Figure 1 displays the decrease in the rate of motility with the applied force as described above (□) and the force corrected for the viscous drag (▲). The total viscous drag, $F_d = f v$, may be calculated for a demembrated sperm by adding the drag on a sphere, $f_s = 6\pi\eta R$ (radius, $R = 1.0 \mu\text{m}$, $\eta = 0.04$ poise), to the drag on a thin rigid rod, $f_r = 4\pi\eta a/(\ln(2a/b) - 0.5)$ ($a = 20 \mu\text{m}$, $b = 0.1 \mu\text{m}$, where a and b are the rod dimensions, $\eta = 0.04$ poise), and allowing for the increased viscous effects of an object near a surface¹⁷. Because these viscous drag forces

stained SDS gels (insets on right). a, Velocity of motility decreases linearly as the opposing force increases under 'high specific functional density' conditions ($0.5 \mu\text{g}$ applied kinesin produced 1 functional motor per μm^2). These conditions were achieved by precoating the coverslip with 5 mg ml^{-1} casein as described in ref. 14. Each point is the average of 10–15 measurements, and the dotted line represents the force at which motility stopped. b, The velocity of movement is also decreased by forces applied in the direction of kinesin-driven motility (other conditions as in a). c, Velocity of motility decreases in a curvilinear fashion as the opposing force increases under 'low specific functional density' conditions. Here, $3 \mu\text{g}$ kinesin produced 1 functional motor per μm^2 . These conditions were achieved by mixing only $100 \mu\text{g ml}^{-1}$ casein with the kinesin and Mg-ATP before adsorption on the coverslip. d, Velocity of motility also decreases in a curvilinear fashion when forces are applied in the same direction as the motility (other conditions same as c).

METHODS. a, Kinesin was purified from sea urchin eggs by modifications of published procedures^{2,15}. Briefly, 200 ml cytosol was treated with hexokinase and glucose, centrifuged to sediment actomyosin, then supplemented with taxol ($15 \mu\text{M}$), GTP (1 mM) and sodium tripolyphosphate (10 mM). The resulting kinesin-microtubule complexes were sedimented through a sucrose cushion, washed in Mg^{2+} -free buffer, and kinesin was eluted by centrifugation in Mg-ATP. The eluate was filtered through a $0.45 \mu\text{m}$ syringe filter, concentrated to 2 ml in a centrprep-30 and fractionated on a $1.6 \times 90 \text{ cm}$ column of Biogel A1.5M. Kinesin was pooled and bound to salt-stripped microtubules in the presence of AMP-PNP (adenyl-5'-yl imidodiphosphate), pelleted, then eluted using Mg-ATP, and finally sedimented on a 5-ml 5–20% sucrose density gradient (300,000g, 8.5 h). Highly purified kinesin ($100\text{--}200 \mu\text{g}$) was obtained and assayed within 48 h. b, We assayed the kinesin-driven movement of demembrated sea urchin sperm over a coverslip mounted into a rotating microscope stage of a Zeiss video microscope equipped with a $\times 32$ or $\times 100$ objective (numerical aperture ~ 1.0). The stroboscopic light source was triggered to flash in synchrony with the rotating stage by a photonic sensor control unit. Motion of the sperm head was recorded on a video recorder, and the data captured by a frame grabber were processed using the public domain software NIH IMAGE v.1.45. Velocities were measured from distances measured in time-sequenced captured frames. We excluded sperm heads that were moving at rates obviously inconsistent with kinesin-driven motility, and were thought instead to be due to convection in the cell.

Replication timing can be conveniently measured by *in situ* hybridization to interphase nuclei⁹; nuclei in which the gene has not yet replicated show two single hybridization dots, whereas those in which the gene has replicated reveal two sets of doublets. Thus in an unsynchronized cell-cycle population, the relative

time of replication can be easily measured as the higher the percentage of doublet nuclei, the earlier the gene replicates in S-phase. In general, most gene loci replicate on the two homologous chromosomes in a fairly synchronous manner⁹, with only about 10% of the nuclei showing an asynchronous single-double

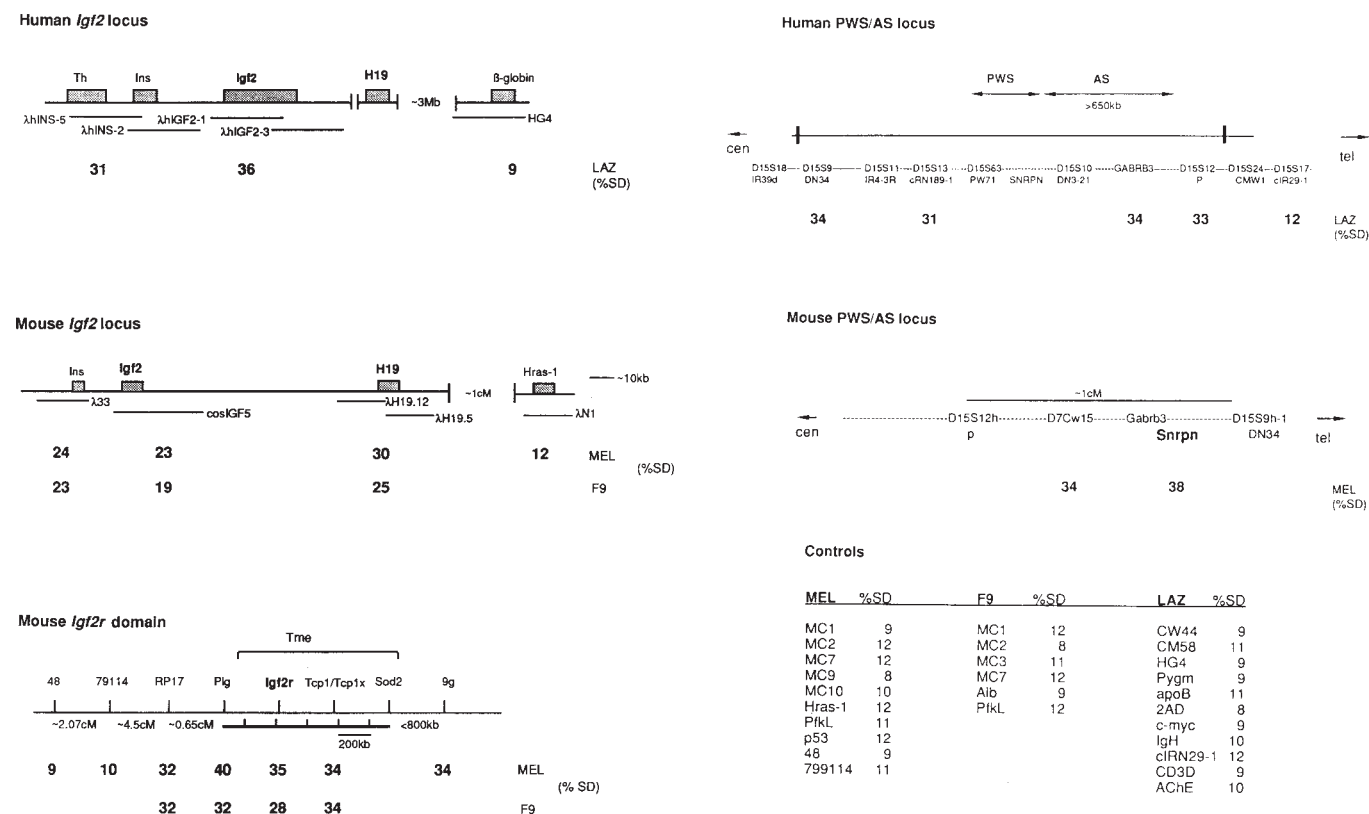


FIG. 1 Asynchronous replication of imprinted genes. Genomic maps of the mouse *Igf2r* region (chromosome 17)², the *Igf2* regions in mouse (distal chromosome 7)^{3,23,31,32} and man (11p15.5)^{26,33,34}, and the human (15q11-13) and mouse (central chromosome 7) PWS/AS regions^{5,30,34} are shown together with the approximate position of phage and cosmid clones used for hybridization analysis (only the *Igf2* maps are drawn to approximate scale). In the human *Igf2* domain, the distance between *Igf2* and *H19* is <200 kb²³. In the PWS/AS locus, the distance between *DN34* and *P* is about 3–4 Mb in man and about 1 cM in the mouse³⁰. The minimal shortest region of deletion overlap in PWS (exact size not known) and AS is also shown^{10,11,34}; thick bars demarcate the outer boundaries of the deletions found in 95% of PWS/AS patients with deletions³⁴. Genes known to be imprinted are indicated in bold type. Fluorescence *in situ* hybridization was done on nuclei from mouse F9 teratocarcinoma cells, mouse erythroleukaemia cells (MEL), or a human lymphoblastic cell line (LAZ; provided by H. Lazarus). For each probe, over 100 nuclei were screened and the number of cells containing singlet/doublet hybridization patterns was recorded. In the case of F9, nuclei from all phases of the cell cycle were scored, but for MEL and human lymphoblast cells bromodeoxyuridine (BrdU) incorporation was used to identify and assay S-phase cells exclusively. The percentage of cells with a single-double signal (%SD) is shown on the map below each probe. In addition to the probes shown, %SD was measured for *Rb* (41%), *WT* (41%), *Abl* (43%) and *Bcr* (40%) in LAZ cells. For each probe, the percentage of cells with two single dots (%SS) can be used as an indication of the relative time of replication in the cell cycle. In the case of asynchronous loci, the percentage of cells with two single dots represents the replication time of the first allele (Fig. 2). A sample of these data is given as follows: in MEL cells, *Igf2* (35%), *Insulin* gene (41%), *H19* (36%), *Igf2r* (24%), *Tcp1* (28%), *Plg* (25%), *RP17* (29%), *79114* (33%), *C48* (46%), *Snrpn* (50%), *D7Cw15* (56%), *H-ras-1* (36%), *PfK-L* (80%). In LAZ cells, *DN34* (58%), *cRN189-1* (59%), *GABRB3* (54%), *cIRN29-1* (23%), *apoB* gene (77%), *acetylcholine esterase* (AChE) (71%), *IgH* (77%), *CD3D* (30%), *Pygm* (28%). Abbreviations: T-associated maternal effect, Tme; megabase, Mb; centimorgan, cM; centromere, cen; telomere, tel.

METHODS. The labelling of genomic phage and cosmid probes with biotin-11-dUTP (Sigma) and the procedures for *in situ* hybridization and detection have been described⁹. For some cultures, S-phase cells were identified by labelling with 10^{-5} M BrdU for one hour before harvesting. After the hybridization, washing and blocking steps, slides were incubated with anti-BrdU antibody (Beckton-Dickinson; 1:10 dilution) at 37 °C for 30 min, and then washed 3 times for 5 min each in $4 \times$ SSC, 0.1% Tween 20 at 42 °C. Detection was completed by incubation with a rhodamine-conjugated goat anti-mouse IgG (H + L) (Jackson Immunochemicals; 1:100) for 30 min at 37 °C together with the avidin-FITC used to visualize the hybridization dots. The following cosmids, phages or phage pairs were used in this study: λ hIGF2-1, λ hIGF2-3, λ hINS-5 and λ hINS-2 (G. Bell); λ alb-5 and λ alb-6 (mouse albumin), λ H19.5 and λ H19.12 (S. Tilghman); λ 33 (L. Villa-Komaroff); cos940 (*Igf2r*), cos 5b (plg), cos 9g, cos RP17, LA2 (tcp1 and tcp1x), 79114, and the human *Igf2r* cosmid (cosSIV) (D. Barlow, N. Schweifer, R. Stoger); cos4.7 (C48 from D15LEH48 locus) (B. G. Herrmann and H. Lehrach); cosIGF-5 (P. Rotwein); a mouse phosphofructokinase (liver type) (PfK-L) cosmid clone (Y. Groner); CW44 and CM58 from the human cystic fibrosis gene region (L. Tsui); HG4 and TK-28 from the human β -globin domain (F. Grosveld); Pygm and GABRB3 (GABA β 3 receptor) (D. C. Ward); *Snrpn* (C. I. Brannan, N. G. Copeland, N. A. Jenkins); λ 2AD (MA β 1B) phage (L. Kunkel); the human apoB (apolipoprotein B) plasmid, 12D1 (B. Levy-Wilson); the mouse *H-ras-1*, λ N1 (A. Balmain); human CD3D cosmid (M. Baniash); human AChE phage (H. Soreq); human *c-myc* cosmid (P. Leder); human *P* (pink-eye dilution) locus (two phages λ G1 and λ G47 from DN10) (R. A. Spritz); mouse *D7Cw15* (E. M. Rinchik and T. Magnuson); cRN189-1 and cIRN29-1, a cosmid clone isolated using probe pIR29-1, itself derived from a 15q11-q13 library; DN34 (3-kb cDNA and 6-kb genomic clone, with 0.5-kb overlap)²⁰; a 16-kb mouse p53 plasmid (M. Oren); Wilm's tumour cosmid WT4.3 (B. Williams), and a series of phages from the retinoblastoma gene (Y. Fung). Pre-labelled cosmid probes for *Abl* and *Bcr* were obtained from ONCOR (Gaithersburg, MD), MC1, MC2, MC3, MC7, MC9 and MC10 are random clones derived from a mouse genomic cosmid library (Clontech).

hybridization pattern (Fig. 1). In contrast, for all of the mouse imprinted genes (*Igf2*, *Igf2r*, *H19* and *Snrpn*), approximately 25–35% of the nuclei from several different cell lines show asynchronous replication, and the same is true for a number of other flanking sequences at each of these loci. A similar pattern of differential replication was also observed in human cells for the *Igf2* and *Igf2r* genes, whose imprinting status has not yet been ascertained, and for markers near the *Snrpn* homologue^{5,10} in the human chromosome region 15q11–q13, which has been implicated in the phenotypically imprinted Prader-Willi and Angelman syndromes¹¹. Furthermore, other genes located in genomic regions that undergo parentally imprinted deletions (genes affected in retinoblastoma and Wilms' tumour)¹² or translocations (*Abl* and *Bcr*)¹³ also have a high percentage of single-double hybridization signals in nuclei from normal lymphoblasts (see legend to Fig. 1). When taken together, these data strongly suggest that asynchronous replication represents a general characteristic of imprinted gene regions in both embryonic and somatic cells.

To determine whether replication timing of imprinted gene regions is allele-specific, we took advantage of cell lines with uniparental copies of these loci. For the *Igf2r* gene, for example, two hemizygous fibroblast cell cultures containing either the paternally or maternally derived allele were established from 10–12-day T^hp mouse embryos¹⁴. *In situ* hybridization analysis of the single *Igf2r* gene in these cells reveals that the paternal allele consistently replicates earlier than the maternal homologue (Fig.

2a). A similar approach was used to assess the allele-specific replication timing of 4 probes in the human 15q11–q13 chromosome region (Fig. 2b). Lymphoblast cells from Prader-Willi syndrome (PWS) patients with a paternal deletion in this region showed that this entire domain replicates uniformly late on the maternal allele, and a reciprocal experiment on Angelman syndrome (AS) cells confirmed that the paternal allele replicates earlier in the cell cycle. Early replication of the paternal allele was also seen in cells from PWS and AS patients with monoallelic disomies of chromosome 15 (Fig. 2b). Whereas these lymphoblasts have two copies of each probe, they are both derived from the same parent and therefore replicate synchronously in S phase.

In a more direct proof of allele-specific replication at the PWS/AS locus, we took advantage of a naturally occurring centromeric satellite polymorphism to identify the parental origin of each chromosome 15 in normal lymphocytes. In one individual (Y.C.), the late-replicating allele was almost exclusively associated with the maternally derived distinctively enlarged satellite region (Fig. 3a). In his mother (Z.C.), however, this same polymorphic chromosome was paternally derived and underwent early replication at the PWS locus (Fig. 3b). Using a similar approach, we analysed allele-specific replication of the human *Igf2* locus in a thalassaemia cell line which contains a 30-kilobase maternal deletion in the nearly β -globin domain¹⁵. Using a globin probe to identify the paternal chromosome, it was possible to show that the adjacent *Igf2* gene consistently replicates earlier

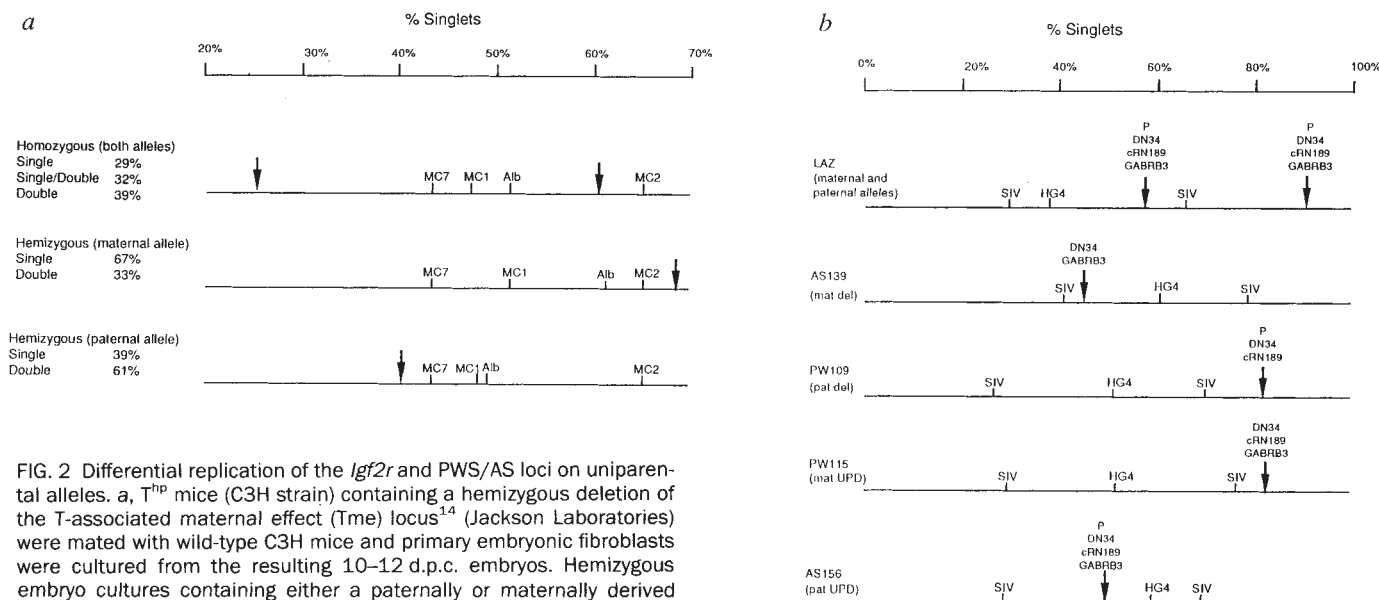


FIG. 2 Differential replication of the *Igf2r* and PWS/AS loci on uniparental alleles. **a**, T^hp mice (C3H strain) containing a hemizygous deletion of the T-associated maternal effect (Tme) locus¹⁴ (Jackson Laboratories) were mated with wild-type C3H mice and primary embryonic fibroblasts were cultured from resulting 10–12 d.p.c. embryos. Hemizygous embryo cultures containing either a paternally or maternally derived *Igf2r* gene domain were distinguished from homozygous embryos using *in situ* hybridization to visualize the number of *Igf2r* alleles per nucleus. These cells were then analysed for replication timing by fluorescence *in situ* hybridization using the *Igf2r* probe on BrdU-positive S-phase nuclei. For homozygous cells we assume that each parental allele replicates at a different specific time in S. In cells with 2 single dots, neither allele has yet replicated and this percentage thus represents the point at which the first allele undergoes DNA synthesis (29%). In nuclei with a single-double hybridization pattern, one allele has replicated, but the second has not, and the addition of this percentage (32%) indicates the point of replication of the second allele (61%). As hemizygous cells contain only one allele, the time of replication is defined by the percentage of cells showing a single dot. All three cell lines were analysed by FACS and seemed to have similar cell-cycle profiles, which were confirmed by comparing the time of *Igf2r* replication to that of four independent cell-cycle markers. The raw data obtained for the homozygous non-deleted cells is shown. For the hemizygous cell cultures, the percentage of single hybridization dots has been normalized to those in the homozygous line by fixing the distance between the two extreme markers (MC2 and MC7). The bold arrows represent the positions of replication of each *Igf2r* allele. It should be noted that these numbers do not reflect

the exact point in the cell cycle at which these genes replicate, because the data do not take into consideration the non-equal distribution of cells in different fractions of S phase. On the other hand, these measurements do provide an accurate indication of their relative replication times. Using northern blot analysis, we have shown that the *Igf2r* gene is transcribed exclusively from the maternal allele in embryonic fibroblasts (data not shown). **b**, Similar analysis of the human PWS/AS locus using BrdU-positive S-phase nuclei from normal diallelic lymphoblasts (LAZ) and patient-derived lines, PW109 (paternally deleted) and AS139 (maternally deleted). Cell-cycle replication-time markers included cosHG4 (β -globin) plus the early- and late-replicating alleles of cosSIV(*Igf2*). Average replication times for four probes in the PWS/AS region (GABRB3, DN34, cRN189-1 and P) are indicated by bold arrows (variability between these probes was less than 5% in each cell type). Cell lines containing either a maternal (PW115) or paternal (AS156) chromosome 15 uniparental disomy (UPD) were analysed in the same way. In contrast to normal cells, the two alleles replicate synchronously in both cell types ($\sim 8\%$ SD for probes DN34 and GABRB3).

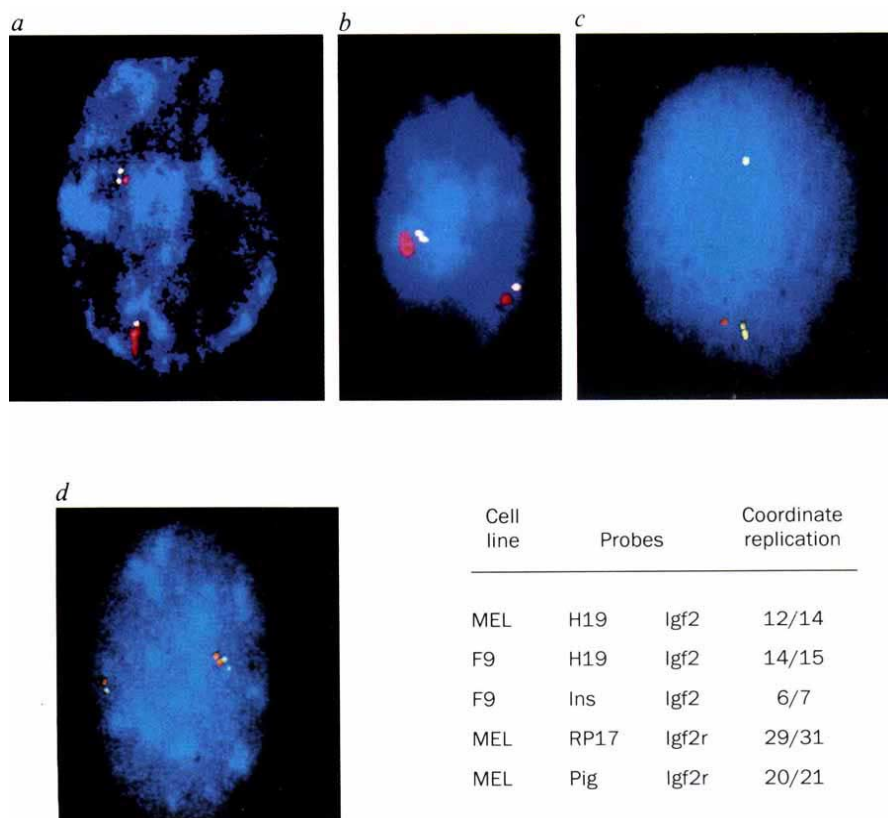
than the maternal allele in the same cell (Fig. 3c). In contrast to these results, the low level of asynchronous replication seen for control probes ($\sim 10\%$) is completely random with respect to the parental alleles (see Fig. 3c legend). Thus, allele-specific replication timing appears to be a special characteristic of imprinted gene regions.

The entire genome is known to be organized into a series of 1–2 Mb replication bands^{9,16} whose specific timing in S phase is probably controlled at the molecular level by long range cis acting sequences¹⁷. Imprinted genes, however, are apparently imbedded in novel replication timing units which can be differentially regulated on the two parental chromosomes. These regions may be the size of replication bands and are bordered by normal synchronously replicating DNA markers (Fig. 1). Within these regions, all of the sites on each allele replicate coordinately at the same time in S (see Fig. 1 legend). This was shown directly for the *Igf2* and *Igf2r* regions by analysing individual nuclei showing a single-double replication pattern for

pairs of genes within each locus (Fig. 3d). In addition, markers within the PWS/AS region all replicate simultaneously on individual parental chromosomes (Fig. 2b), and for this locus, differential replication bands have actually been observed morphologically¹⁸. Parent-specific replication banding is also characteristic of the imprinted X chromosome in extraembryonic tissues in the mouse¹⁹. At these loci there is no straightforward correlation between the time of replication of a particular gene allele and its potential activity state in imprinted tissues (see Fig. 3c legend). This is best seen for the two adjacent *Igf2* and *H19* genes that are coordinately replicating on each parental allele (Fig. 3c), despite being imprinted in opposite directions. Thus, rather than being gene-specific, the determination of replication timing is regional and strikingly parent-specific, with the paternal DNA being early replicating in every case¹⁹ (Figs 2 and 3).

These results suggest the existence of chromosomally imprinted domains characterized structurally by allele-specific replication and methylation^{20,21} patterns and functionally by

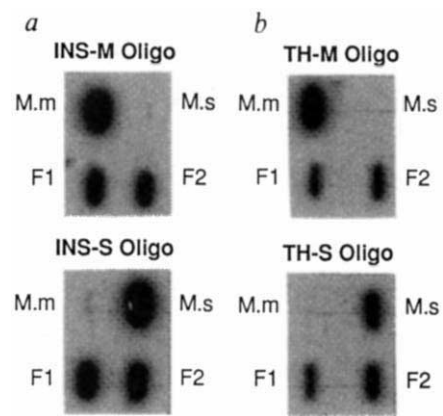
FIG. 3 Allelic and coordinate replication timing in imprinted gene regions. *a* and *b*, Individual Y.C. is heterozygous for a $15s^+$ chromosome³⁵, shown by cytogenetic analysis to be maternally derived. Primary lymphocyte cultures from Y.C. and his mother (Z.C.) were prepared from peripheral blood and after 72 h cells were collected and hybridized *in situ* with a digoxigenin-labelled *GABRB3* probe (detection with anti-DIG FITC) and a biotinylated probe of chromosome 15 satellite DNA (ONCOR) (detected with avidin-rhodamine). The microscopic image was photographed and pseudocoloured as described³⁶. In each cell, hybridization to the $15s^+$ satellite locus (red) is more extensive than that of its homologue and this can be used to identify the parental origin of the nearby $15q11-q13$ region. Only cells showing a single-double pattern for the *GABRB3* gene (yellow) were scored. In YC, 21/23 showed the single dot associated with the maternally derived s^+ chromosome. For Z.C. nuclei, 24/25 showed the double dot associated with the paternally derived $15s^+$ chromosome. *c*, Nuclei from the human Hispanic $\lambda\delta\beta$ thalassaemia cell line were hybridized to a digoxigenin (Boehringer)-labelled HG4 globin cosmid clone which is specific for a region 5' to the ϵ gene deleted on the maternal chromosome in these cells¹⁵. Hybridization was detected with a rhodamine-conjugated anti-DIG antibody (red). Biotin-labelled *Igf2* was hybridized simultaneously to the same preparation and detected with avidin-FITC (yellow). Out of 51 nuclei containing a single-double *Igf2* signal, 40 (80%) had the doublet associated with the rhodamine-labelled paternal allele. Controls: In contrast to DNA located within imprinted regions, control probes have single-double signals in only $\sim 10\%$ of nuclei, and this is probably attributable to technical inaccuracies of the *in situ* hybridization technique^{9,36}. In double-label experiments it was shown that this low level of asynchronous replication is not allele-specific. Thus for cIR29-1 (FITC), 18/34 nuclei had the double hybridization signal associated with the 15^+ (RITC) chromosome in Y.C. cells. TK-28 (FITC), a cosmid probe located in the human β -globin domain, had the double dot on the paternal chromosome in 13/28 single-double nuclei when measured in the Hispanic $\gamma\delta\beta$ thalassaemia cell line using the nearby HG4 cosmid (RITC) to distinguish between parental alleles. In mouse embryonic fibroblasts carrying only one *Igf2r* gene, the external probe C48 (FITC) had the double hybridization signal on the maternal allele (detected using *Igf2r* probe labelled with RITC) in 11/22 nuclei. *d*, To test coordinate replication timing on adjacent genes, *in situ* hybridization was carried out using one probe labelled with biotin (detection by avidin-FITC) and another labelled with digoxigenin (detection by anti-DIG rhodamine) on either mouse F9 or MEL cells.



Only cells with a single-double replication pattern for both probes were scored. An example of one such nucleus probed simultaneously with *Igf2* and *H19* is shown. If both genes replicate coordinately, they should be in the same replication state at each allele. On the other hand, if *Igf2* is early-replicating on one allele, while the adjacent *H19* is late-replicating, singlets of one probe should be associated with doublets of the other. The number of nuclei showing coordinate replication out of the total number scored for several different gene pairs is presented in the table. Nuclei that did not have the single-double hybridization pattern were also analysed, and here too neighbouring probes were coordinately replicating in about 80–90% of the cells. Thus, each parental locus clearly replicates within a defined window of S, but there could be small differences in the timing of individual gene sequences within each domain. Following is a summary of the state of expression and replication timing of known murine imprinted genes. *Igf2* (ref. 4) and *Snrpn* (ref. 5) are maternally inactive and associated with the late-replicating chromosome. *Igf2r* (ref. 2), *H19* (ref. 3) and genes on the X chromosome in extraembryonic tissues⁸ are paternally imprinted and associated with the early-replicating chromosome¹⁹.

FIG. 4 Diallelic expression of mouse insulin II and tyrosine hydroxylase. **a**, RNA from pancreas of *M. musculus* (M.m.), *M. spretus* (M.s.), and F₁ or F₂ interspecific hybrids was isolated and amplified by insulin II-specific polymerase chain reaction (PCR). Equal amounts (~50 ng) of the resulting DNA was then analysed by slot-blot hybridization using species-specific *M. musculus* (INS-M) and *M. spretus* (INS-S) oligo-nucleotide probes. Similar results were obtained using RNA from 18-d embryonic pancreas. **b**, Analysis of tyrosine hydroxylase (TH) RNA from adult brain tissue. Identical results were obtained with adult adrenal mRNA. Using *M. musculus* (TH-M) and *M. spretus* (TH-S) specific oligo-nucleotide probes.

METHODS. F₁ hybrids were obtained by crosses between *M. spretus* males and *M. musculus* females. To generate mice containing a maternally derived *M. spretus* insulin II gene, F₁ females were backcrossed to *M. musculus* males and the resulting animals or embryos screened for the presence of the *M. spretus* allele using a specific oligonucleotide probe (see below). Allele-specific RNA analysis was possible because of differences between the genes from *M. musculus* and *M. spretus*. A PCR fragment from *M. spretus* insulin II RNA was synthesized using primers that distinguish this product from the insulin I gene (5'AGCCCTAAGTGATCCGCTAC3' and 5'GAACCACAAAGGTGCTGC3'). Sequencing of the PCR product indicated the presence of a polymorphism at nucleotide -12, and oligonucleotide probes specific for *M. musculus* (5'AGGTATTGTTTCAACA3') and *M. spretus* (5'AGGTGATTGTTTCAACA3') were then prepared from this region. A similar strategy was used to detect species-specific transcripts of the



tyrosine hydroxylase gene. In this case the PCR primers were 5'TTC-TGGCCAGTCTGGCCTTC3' and 5'TCCTCTGACAGGGAGTGACAG3', and a polymorphism was found at nucleotide 1,101 of the mRNA. *M. spretus* (5'AACTGCTCCACGGTGTACT3') and *M. musculus* (5'AACTGCTCCACGGTGTACT3') specific oligonucleotide probes were end-labelled and analysed by slot blotting as described³⁷.

their content of multiple imprinted genes^{8,11,22,23}. It was therefore of interest to investigate whether other genes in these regions may also be transcribed in an allele-specific manner. To this end we took advantage of *Mus musculus*-*Mus spretus* interspecific hybrids to assay allele-specific transcription of the insulin II and tyrosine hydroxylase genes, which are both located 5' to *Igf2* (ref. 24), in the asynchronous replicating region. As shown in Fig. 4, both parental alleles are equally transcribed, and the same is probably true of the human insulin gene²⁵, which is also located²⁶ in a differentially replicating region containing at least one known imprinted gene (H19; refs 27, 28). Non-imprinted genes are also included within the asynchronous replicating *Igf2* domain² and the PWS/AS locus in man^{29,30}.

It is clear that the mechanism of imprinting must involve epigenetic signals which are acquired during gametogenesis and then stably transmitted to cells of the developing organism. Although our data show a close correlation between imprinting

of individual genes and their location within differentially replicating regions, the functional relationship between these parameters is not yet clear. On the one hand, allele-specific replication could simply be a secondary regional by-product of imprinting at a specific gene locus. Alternatively, establishment of a structural replication imprint at individual chromosome regions may actually represent an early event that precedes the expressional imprinting of specific genes in these domains. Our data support this second possibility, because differential replication is already present in F9 cells and embryonic fibroblasts and is stably inherited in a variety of somatic cell types, including those that do not show an imprinted phenotype. This thus represents a developmentally conserved structural imprint which, together with local *trans*-acting factors, could play a part in the control of allele-specific transcription.

Note added in proof: It has recently been shown that the human *Igf2* gene is imprinted^{38,39}. □

Received 30 March; accepted 13 May 1993.

1. Solter, D. *Rev. Genet.* **22**, 127-146 (1988).
2. Barlow, D. P., Stoger, R., Herrmann, B. G., Saito, K. & Schweifer, N. *Nature* **349**, 84-87 (1991).
3. Bartolomei, M. S., Zemel, S. & Tilghman, S. M. *Nature* **351**, 153-155 (1991).
4. DeChiara, T. M., Robertson, E. J. & Efstratiadis, A. *Cell* **64**, 849-859 (1991).
5. Leff, S. E. *et al. Nature Genet.* **2**, 1-5 (1992).
6. Goldman, M. A., Holmquist, G. P., Gray, M. C., Caston, L. A. & Nag, A. *Science* **224**, 686-692 (1984).
7. Holmquist, G. P. *Am. J. hum. Genet.* **40**, 151-173 (1987).
8. Takagi, N. & Oshimura, M. *Expl. Cell Res.* **78**, 127-135 (1973).
9. Selig, S., Okumura, K., Ward, D. C. & Cedar, H. *EMBO J.* **11**, 1217-1225 (1992).
10. Ozcelik, T. *et al. Nature Genet.* **2**, 2-6 (1992).
11. Nicholls, R. D., Rinchik, E. M. & Driscoll, D. J. in *Seminars in Developmental Biology* (eds Surani, M. A. & Reik, W.) 139-152 (Academic, London, 1992).
12. Reik, W. & Surani, M. A. *Nature* **338**, 112-113 (1989).
13. Haas, O. A., Argyriou-Tirita, A. & Lion, T. *Nature* **359**, 414-416 (1992).
14. Johnson, D. R. *Genetics* **67**, 795-804 (1974).
15. Driscoll, C., Dobkin, C. S. & Alter, B. P. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7470-7474 (1989).
16. Drouin, R., Lemieux, N. & Richer, C. L. *Chromosome* **99**, 272-280 (1990).
17. Forrester, W. C. *et al. Genes Dev.* **4**, 1637-1650 (1990).
18. Izumikawa, Y., Naritomi, K. & Hirayama, K. *Hum. Genet.* **87**, 1-5 (1991).
19. Sugawara, O., Takagi, N. & Sasaki, M. *Cytogenet. Cell Genet.* **39**, 210-219 (1985).
20. Driscoll, D. J. *et al. Genomics* **13**, 917-924 (1992).
21. Dittich, B. *et al. Hum. Genet.* **90**, 313-315 (1992).
22. Forejt, J. & Gregorova, S. *Cell* **70**, 443-450 (1992).

23. Zemel, S., Bartolomei, M. S. & Tilghman, S. M. *Nature Genet.* **2**, 61-65 (1992).
24. Saunders, A. M. & Seldin, M. F. *Genomics* **6**, 525-535 (1990).
25. Shoelson, S. *et al. Nature* **302**, 540-543 (1983).
26. Junien, C. & van Heyningen, V. *Cytogenet. Cell Genet.* **58**, 459-554 (1991).
27. Zhang, Y. & Tycko, B. *Nature Genet.* **1**, 40-44 (1992).
28. Rachmilewitz, J. *et al. FEBS Lett.* **309**, 25-28 (1992).
29. Rinchik, E. M. *et al. Nature* **361**, 72-76 (1993).
30. Nicholls, R. D. *et al. Proc. natn. Acad. Sci. U.S.A.* **90**, 2050-2054 (1993).
31. Rotwein, P. & Hall, L. *DNA Cell Biol.* **9**, 725-735 (1990).
32. Rinchik, E. *Mammalian Genome* **3**, S104-120 (1992).
33. Bell, G. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 6450-6454 (1982).
34. Kuwano, A. *et al. Hum. molec. Genet.* **1**, 417-425 (1992).
35. Chen, T. R., Kao, M. L., Marks, J. & Chen, Y. Y. *Am. J. med. Genet.* **9**, 61-66 (1981).
36. Lichter, P. *et al. Science* **247**, 64-69 (1990).
37. Conner, B. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 278-282 (1983).
38. Ranier, S. *et al. Nature* **362**, 747-749 (1993).
39. Ogawa, O. *et al. Nature* **362**, 749-751 (1993).

ACKNOWLEDGEMENTS. S.S. and D.K. contributed equally to this work. We thank all those who provided clones for this study (see Fig. 1); M. Kotler for embryonic cells; Y. Cedar, Z. Cedar and D. Abeliovitch for assistance; M. Groudine and T. Papayannopoulou for the Hispanic $\gamma\delta\beta$ thalassaemia cell line and H. Lazarus for the LAZ cell line. This research was supported by grants from the NIH (H.C.), the Israel Cancer Research Fund (H.C.), the Forchheimer Foundation (H.C.), and the American Cancer Society (R.D.N. and D.J.D.). R.D.N. is a Pew Scholar in the Biomedical Sciences.

Design, synthesis and biological activity of protaxols

K. C. Nicolaou*, C. Riemer*, M. A. Kerr*,
D. Rideout† & W. Wrasidlo‡

* Department of Chemistry, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, California 92037, and Department of Chemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA

† Department of Molecular Biology, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, California 92037, USA

‡ Drug Discovery Unit, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, California 92037, USA

TAXOL¹⁻⁶ is a product isolated from the Pacific yew tree (*Taxus brevifolia*) and is a potent microtubule-stabilizing agent which has recently been approved for treatment of otherwise intractable ovarian cancer. Despite taxol's therapeutic promise, its aqueous insolubility ($<0.004 \text{ mg ml}^{-1}$) hampers its clinical application. Here we report the design, synthesis and biological activity of a series of taxol-releasing compounds (protaxols) with improved pharmacological properties. These prodrugs were designed to increase their aqueous solubility and allow for taxol release under basic or physiological conditions. We demonstrate the stability of these prodrugs at $\text{pH} \leq 7$ and their ability to release taxol in a basic medium. Taxol-like microtubule-stabilizing activity⁷⁻⁹ appears after the release of taxol. *In vitro* these prodrugs have cytotoxic properties against tumour cell lines comparable to those of taxol; moreover, human plasma catalyses the release of active taxol. These protaxols have greater potential as anticancer agents than the parent compounds taxol and taxotere (Fig. 1a).

Structure-activity relationships and variations in solubility of taxol have been investigated through substitution at the C-2' hydroxyl group¹⁰⁻¹³, but new taxols are needed with improved pharmacological properties. We therefore designed two types of structure (I and II in Fig. 1b). As the C-2' hydroxyl group is the most convenient site to attach designed functional domains and because blocking this position results in loss of activity¹⁰⁻¹³, we selected groups fulfilling the following criteria: (1) the functional group should initiate its own cleavage from the conjugate and generate taxol *in situ* upon being appropriately activated; (2) the newly introduced group should increase the water-solubility of the compound. Also, as certain drug-resistant tumour cells have microenvironments of basic pH^{14} , we focused on base-labile moieties as activating groups. Both types of attached group in structures I and II should be cleaved in basic conditions or by enzymatic reaction (Fig. 1b).

On the basis of these considerations, the taxol derivatives 3-12 shown in Table 1a were designed and synthesized. Carbonates were prepared by derivatization with an appropriate (2-thio-aryl)ethyl chloroformate; esters were formed by reaction with a suitable anhydride. Further manipulations yielded additional compounds for testing.

The aqueous solubilities of the synthesized protaxols were determined in D_2O by ^1H NMR spectroscopy, which revealed that solubilities were considerably higher (Table 1a) than taxol in the case of type II compounds, whereas compounds of type I (Fig. 1b) showed little improvement in water solubility.

All of our designed compounds were very stable in aprotic organic solvents and in the solid state, but in aqueous media some of these compounds had a tendency to revert slowly to taxol even at $\text{pH} 7.0$. Derivatives in which the C-7 hydroxyl group were acetylated were extremely unstable in aqueous media, losing the C-7 substituent rapidly. The stability of the synthesized compounds was tested at various pHs and temperatures. Figure 2a summarizes results obtained with compound 5,

in which the accelerating effect of higher temperatures and pH on taxol release was demonstrated. Table 1a includes the half-lives of six of these compounds at pH 7.5 and 9. As seen from these data, for compounds of type I the rate of taxol release increases with the electron-withdrawing ability of the aryl substituents, whereas for type II derivatives, the rate of release increases with the electron-withdrawing nature of the linking heteroatom. This illustrates the capacity for fine-tuning the rate of release of taxol according to a desired set of conditions using simple chemical principles.

Significantly, incubation of compound 5 in human plasma at 37°C accelerates the release of taxol (half-life $t_{1/2} \sim 100 \text{ min}$) when compared to taxol release in aqueous media at the same pH and temperature (Fig. 2b). This indicates that taxol release from this compound (5) is being assisted by factors present in human plasma and suggests that it may be effective *in vivo*.

Tubulin polymerization to microtubules is promoted by GTP, whereas CaCl_2 causes depolymerization of microtubules back to tubulin (Fig. 3a). Taxol allows and promotes this type of microtubule assembly and, furthermore, it stabilizes microtubules against CaCl_2 -induced depolymerization (Fig. 3b). Protaxol 5 was tested for its ability to stabilize microtubules formed from tubulin as a result of the action of GTP. This agent failed to prevent CaCl_2 -induced disassembly of microtubules at the initial stages of the experiment (Fig. 3c), but showed increased potency with prolonged exposure times, as would be expected from the slow release of taxol, reaching a level comparable to that of taxol when conversion to taxol was essentially complete (Fig. 3d). Similar results were obtained with several other protaxols (Table 1). These findings are in agreement with the previously reported loss of activity resulting from blocking the C-2' hydroxyl group¹⁰⁻¹³.

The protaxols were tested against a broad range of cell lines, including the multiple-drug-resistant ovarian cells OVCAR-3, and lung (H-322) and leukaemia (MOLT-4) cells, to assess their cytotoxicity, cell-type selectivity, and duration of action. Comparison with taxol itself revealed that not only were potencies similar, but so also were selectivities against various cell lines. This is consistent with a mechanism of action dependent on taxol release, rather than the derivative itself having intrinsic activity. Furthermore, taxol was isolated and its identity confirmed from aqueous solutions of the designed compounds, thereby establishing that the agent was released under conditions appropriate for microtubule assembly and in the cytotoxicity experiments. The most significant findings are shown in Table 1b, which shows that the selectivity of taxol against cancer cells over normal cells

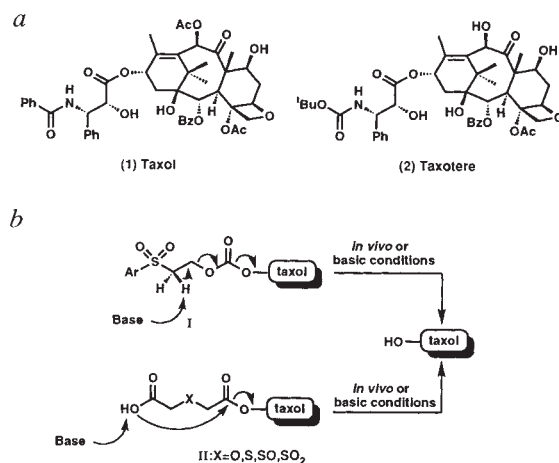
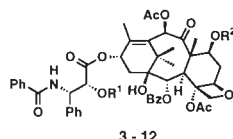


FIG. 1 Molecular structure of taxol and taxotere (a) and mechanistic rationale for the design of protaxols (b) (see text).

TABLE 1 Solubility and stability of protaxols (a) and cytotoxicity of selected compounds (b)

a



	Compound	Water solubility (mg ml ⁻¹) ^k	t _{1/2} ⁿ (pH 7.5)	t _{1/2} ⁿ (pH 9)
1: Taxol (R ¹ =R ² =H)		<<0.10		
3 ^a : R ¹ =	; R ² =H	0.84	>500	110
4 ^b : R ¹ =	; R ² =H	0.35	>500	300
5 ^c : R ¹ =	; R ² =H	1.2 ^m	>500	30
6 ^d : R ¹ =	; R ² =H	—	—	—
7 ^e : R ¹ =	; R ² =H	—	—	—
8 ^f : R ¹ =	; R ² =H	<0.10 ^l	>500	66
9 ^g : R ¹ =	; R ² =H	<0.10 ^l	>500	7
10 ^h : R ¹ =	; R ² =H	<0.10 ^l	≥500	101
11 ⁱ : R ¹ =R ² =	—	0.50	—	—
12 ^j : R ¹ =R ² =	—	0.46	—	—

b

Cytotoxicity of taxol derivatives (IC₅₀[M])^{*}

		Taxol (1)	5	6	7	9	10	11	12
Normal cell lines									
NHDF	Normal human skin	10 ⁻⁶	10 ⁻⁶	10 ⁻⁶	10 ⁻⁶	10 ⁻⁶	10 ⁻⁶	10 ⁻⁶	10 ⁻⁶
RPMI-7666	Normal human PBLs	10 ⁻⁴	10 ⁻⁴	10 ⁻⁴	10 ⁻⁴	10 ⁻⁴	10 ⁻⁴	10 ⁻⁴	10 ⁻⁴
BALB/c 3T3	Normal mouse embryo	10 ⁻⁶	10 ⁻⁵	10 ⁻⁵	10 ⁻⁵	10 ⁻⁵	10 ⁻⁵	10 ⁻⁶	10 ⁻⁵
CHO	Chinese hamster ovary	10 ⁻⁴	10 ⁻⁴	10 ⁻⁴	10 ⁻⁴	10 ⁻⁴	10 ⁻⁴	10 ⁻⁴	10 ⁻⁴
CHO	Normal human mammary	10 ⁻⁴	10 ⁻⁴	10 ⁻⁴	10 ⁻⁴	10 ⁻⁴	10 ⁻⁴	10 ⁻⁴	10 ⁻⁴
Cancer cell lines									
SK-MEL-28	Melanoma	10 ⁻⁵	10 ⁻⁵	10 ⁻⁵	10 ⁻⁵	10 ⁻⁵	10 ⁻⁵	10 ⁻⁵	10 ⁻⁵
CAPAN-1	Pancreatic carcinoma	<10 ⁻⁹	<10 ⁻⁹	10 ⁻⁷	<10 ⁻⁹	10 ⁻⁶	<10 ⁻⁹	<10 ⁻⁹	<10 ⁻⁹
H322	Lung carcinoma	10 ⁻⁹	10 ⁻⁹	10 ⁻⁷	10 ⁻⁹	10 ⁻⁶	10 ⁻⁹	10 ⁻⁹	10 ⁻⁷
MCF-7	Breast carcinoma	10 ⁻⁵	10 ⁻⁵	10 ⁻⁵	10 ⁻⁵	10 ⁻⁵	10 ⁻⁵	10 ⁻⁵	10 ⁻⁵
BT-549	Breast carcinoma	<10 ⁻⁹	<10 ⁻⁹	10 ⁻⁹	<10 ⁻⁹	10 ⁻⁷	10 ⁻⁷	<10 ⁻⁹	<10 ⁻⁹
OVCA-3	Ovarian carcinoma	<10 ⁻⁹	<10 ⁻⁹	10 ⁻⁶	<10 ⁻⁹	10 ⁻⁶	<10 ⁻⁹	10 ⁻⁶	10 ⁻⁸
HT-29	Colon carcinoma	<10 ⁻⁹	<10 ⁻⁹	10 ⁻⁷	<10 ⁻⁹	10 ⁻⁷	<10 ⁻⁹	<10 ⁻⁹	10 ⁻⁷
SIHA	Cervix carcinoma	10 ⁻⁵	10 ⁻⁴	10 ⁻⁴	10 ⁻⁵	10 ⁻⁵	10 ⁻⁵	10 ⁻⁵	10 ⁻⁵
786-0	Renal cell carcinoma	10 ⁻⁷	10 ⁻⁶	10 ⁻⁵	10 ⁻⁵	10 ⁻⁶	10 ⁻⁶	10 ⁻⁶	10 ⁻⁶
PC-3	Prostate carcinoma	10 ⁻⁶	10 ⁻⁶	10 ⁻⁶	10 ⁻⁶	10 ⁻⁶	10 ⁻⁶	10 ⁻⁶	10 ⁻⁶
HL-60	Promyel carcinoma	<10 ⁻⁹	<10 ⁻⁹	10 ⁻⁷	<10 ⁻⁹	10 ⁻⁷	<10 ⁻⁹	<10 ⁻⁹	<10 ⁻⁹
MOLT-4	T-cell leukaemia	<10 ⁻⁹	10 ⁻⁹	10 ⁻⁷	<10 ⁻⁹	10 ⁻⁹	<10 ⁻⁹	<10 ⁻⁹	<10 ⁻⁹
L1210	Mouse leukaemia	10 ⁻⁶	10 ⁻⁵	10 ⁻⁵	10 ⁻⁵	10 ⁻⁶	10 ⁻⁵	10 ⁻⁶	10 ⁻⁶
UCLA-P-3	Lung carcinoma	<10 ⁻⁹	<10 ⁻⁹	<10 ⁻⁹	<10 ⁻⁹	<10 ⁻⁹	<10 ⁻⁹	10 ⁻⁹	10 ⁻⁹

In a, reagents and conditions for the synthesis of compounds 3–12 from taxol (1) are indicated by superscripts as follows: (a) 1.1 equivalents of diglycolic anhydride in pyridine at 25 °C for 2 h (yield 75%); (b) 2.0 equivs thiodiglycolic anhydride in pyridine at 25 °C for 4 h (67%); (c) as (b), then 3.0 equivs oxone, CH₃OH/H₂O, 25 °C, 12 h (65% overall); (d) as (b), then CH₂N₂, 3.0 equivs oxone, CH₃OH/H₂O, 25 °C, 2 h (32% overall); (e) 1.2 equivs (2-thiophenyl)ethyl chloroformate, (C₂H₅)₃N, CH₂Cl₂, 25 °C, 2 h (94%); (f) as (e), then 3.0 equivs oxone, CH₃OH/H₂O, 25 °C (75% overall); (g) 1.2 equivs (2-p-nitrothiophenyl)ethyl chloroformate, (C₂H₅)₃N, CH₂Cl₂, 25 °C, then oxone, (74% overall); (h) as (g), then oxone, then sodium dithionite, C₂H₅OH, H₂O (62% overall); (i) 10.0 equivs diglycolic anhydride pyridine (87%); (j) 10.0 equivs thiodiglycolic anhydride, 4-dimethylaminopyridine (as catalyst), CH₂Cl₂ (29%). Solubilities: (k) Water solubility was determined by combining 2.0 mg sample with 1.0 ml D₂O and sonicating for 15 min. The resulting mixture was centrifuged for 20 min and the supernatant saturated solution was analysed by ¹H NMR spectroscopy using a Bruker AMX-400 instrument and methanol or pyridine as internal standard. The concentration of the compound was determined by comparing several signals with the CH₃ signal of methanol or the NCH signal of pyridine; (l) the detection limit of this method, as determined by diluting samples of compound 5, was found to be 0.10 mg ml⁻¹; (m) higher concentrations of compound 5 produced hydrogels, presumably due to micelle or related aggregate formation; (n) half-life (t_{1/2}) is defined as the time in minutes at which 50% taxol release was observed at 37 °C. For b, the cytotoxicities were determined using the sulphorhodamine B assay method¹⁵.

* IC₅₀[M] is the molarity at which 50% of tumour cell viability was observed after 72 h exposure under standard tissue culture conditions.

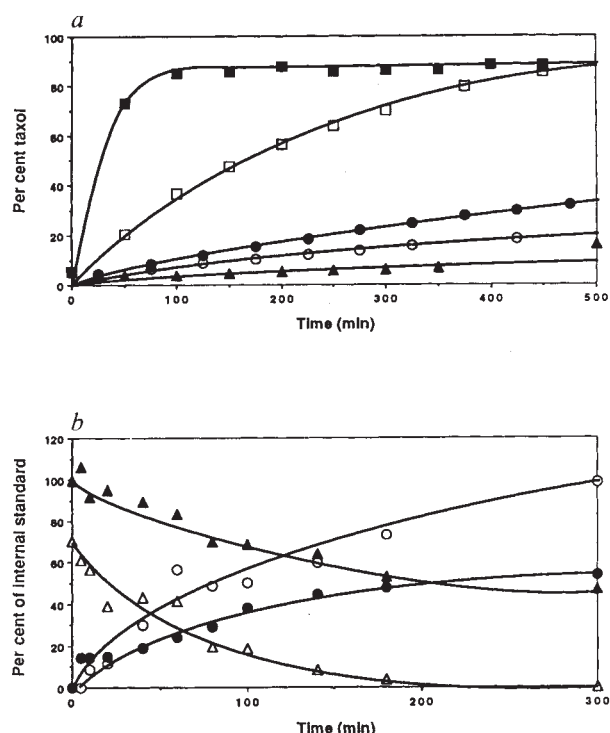


FIG. 2 Kinetics of taxol release from protaxol 5 in aqueous buffer solutions at various pHs and temperatures (a) and in human blood plasma and pH 7.4 buffer solution at 37 °C (b). Although it is quite stable at 20 °C and in pH 7.0 buffer solution, protaxol 5 is converted to taxol in human plasma at 37 °C with a half-life of ~100 min. Conditions in a: $\blacksquare$, pH 7.0, 20 °C; $\square$, pH 7.0, 37 °C; $\bullet$, pH 7.5, 37 °C; $\circ$, pH 8.0, 37 °C; $\blacktriangle$, pH 9.0, 37 °C. In b, $\blacktriangle$, Per cent protaxol 5 remaining in pH 7.4 phosphate buffer; $\bullet$, per cent taxol released in pH 7.4 phosphate buffer; $\triangle$, per cent protaxol 5 remaining in human plasma; $\circ$, per cent taxol released in human plasma.

METHODS. a, Protaxol 5 was dissolved in dimethylsulphoxide (DMSO) at 25 °C and immediately added to the appropriate phosphate buffer solution at the specified temperature. Aliquots were analysed using a Waters Maxima system HPLC instrument equipped with auto-injector (3.9×300 mm C_{18} column equipped with a precolumn; flow rate, 1.5 ml min^{-1} ; eluant, gradient A-B, where A is 80% 100 mM ammonium acetate, buffer pH 6.0, and B is 100% methanol; ultraviolet detector). The per cent taxol released was determined from the relative areas of the peaks corresponding to protaxol 5 and taxol (the conversion was clean, and it was assumed that the two compounds have equal molar extinction coefficients). b, Protaxol 5 and N-cyclohexylbenzamide (as internal standard) were dissolved in acetonitrile and immediately added to human plasma or phosphate buffer at pH 7.4. Samples of protaxol 5 were incubated in each medium for specified times and then extracted with ethyl acetate as described¹⁶. Solvent was removed and the residue dissolved in acetonitrile and analysed by HPLC as already described. The peak areas for compound 5 and taxol were normalized to that of the internal standard using a calibration curve.

is maintained in the designed compounds. Compounds 6 and 9 have lower cytotoxicity after a three-day exposure, which is in line with the decreased rate of taxol release from these conjugates.

Our results demonstrate the application of chemical principles in fine-tuning the molecular design of derivatives of the taxol family for biological and pharmacological purposes. The designed molecules act as prodrugs, releasing the active agent *in situ*. Investigations with tubulin and *in vitro* cytotoxicity experiments have confirmed the unique taxol-like profiles of a number of these compounds. Particularly promising is compound 5,

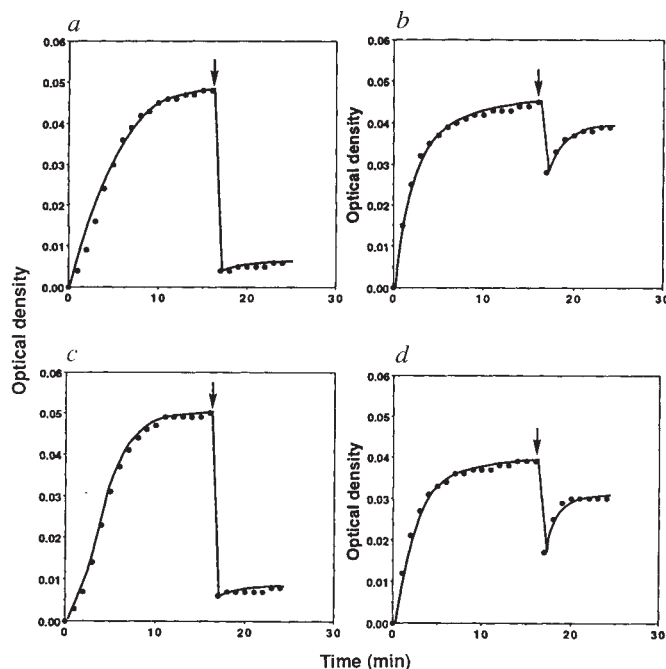


FIG. 3 Tubulin polymerization-depolymerization experiments with taxol and protaxol 5. CaCl_2 -promoted depolymerization of microtubules is suppressed by taxol and protaxol 5 (after taxol release). Experiments were done at 37 °C following a recently developed procedure¹⁷. In each case, 1.0 mM GTP was used to promote initial polymerization of tubulin. a, Control: tubulin (1.0 mg ml^{-1}) alone; CaCl_2 (0.25 mM) added after 16 min caused depolymerization of microtubules. b, Tubulin (1.0 mg ml^{-1}) with taxol (10^{-6} M); CaCl_2 (0.25 mM) added after 16 min did not cause depolymerization. c, Tubulin (1.0 mg ml^{-1}) with protaxol 5 (10^{-6} M); CaCl_2 (0.25 mM) added after 16 min caused depolymerization of microtubules. d, Tubulin (1.0 mg ml^{-1}) with protaxol 5 (10^{-6} M) after release of taxol (as determined by HPLC; essentially a quantitative conversion). The point of addition of CaCl_2 solution is indicated by an arrow.

whose solubility and stability in aqueous media are good, and which is able to release taxol rapidly in human blood plasma. The pharmacology of these compounds needs to be further investigated, as does the design of agents more effective against resistant tumours, which may be achieved by modification of naturally occurring compounds or by total synthesis. $\square$

Received 1 March; accepted 11 May 1993.

- Wani, M. C., Taylor, H. L., Wall, M. E., Coggon, P. & McPhail, A. T. *J. Am. chem. Soc.* **93**, 2325-2327 (1971).
- Nicolaou, K. C., Dai, W.-M. & Guy, R. K. *Angew. Chem. (Int. edn)* (in the press).
- Poitier, P. *Chem. Soc. Rev.* **21**, 113-119 (1992).
- Swindell, C. S. *Org. Chem. Prep. Proc. Int.* **23**, 465-543 (1991).
- Kingston, D. G. I. *Pharmac. Ther.* **52**, 1-34 (1991).
- Rowinsky, E. K. & Donehower, R. C. *Pharmac. Ther.* **52**, 35-84 (1991).
- Schiff, P. B., Fant, J. & Horwitz, S. B. *Nature* **277**, 665-667 (1979).
- De Brabander, M., Geuens, G., Nuydens, R., Willebrords, R. & De Mey, J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5608-5612 (1981).
- Schiff, P. B. & Horwitz, S. B. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1561-1565 (1980).
- Deutsch, H. M. et al. *J. med. Chem.* **32**, 788-792 (1989).
- Swindell, C. S., Krauss, N. E., Horwitz, S. B. & Ringel, I. *J. med. Chem.* **34**, 1176-1184 (1991).
- Zhao, Z. & Kingston, D. G. I. *J. nat. Prod.* **54**, 1607-1611 (1991).
- Mathew, A. E., Mejillano, M. R., Nath, J. P., Himes, R. H. & Stella, V. J. *J. med. Chem.* **35**, 145-151 (1992).
- Keizer, H. G. & Joenje, H. *J. natn. Canc. Inst.* **81**, 706-709 (1989).
- Skehan, P. et al. *J. natn. Canc. Inst.* **82**, 1107-1112 (1990).
- Longnecker, S. M. et al. *Cancer Treat. Rep.* **71**, 53-59 (1987).
- Merlock, R. & Wrasidlo, W. *Analyt. Biochem.* (in the press).

ACKNOWLEDGEMENTS. We thank A. Bustamante, R. Merlock and G. Hum for technical assistance. This work was supported by the NIH, a fellowship from the Alexander von Humboldt Foundation (C.R.) and The Scripps Research Institute.